Esboço de uma teoria semântica da informação[*]

Yehoshua Bar-Hillel; Rudolf Carnap

§1. O problema

Os conceitos de informação e quantidade de informação são distintos. A explicação destes conceitos é tentada apenas na medida em que se aplicam a frases (declarativas) ou, alternativamente, a proposições. A teoria predominante de comunicação (ou transmissão de informações) negligencia deliberadamente a aspectos semânticos da comunicação, ou seja, o significado das mensagens. Esta contenção teórica não é, no entanto, sempre cumprida na prática, e resulta em muitos erros de aplicação. A teoria esboçada aqui é total e abertamente de caráter semântico e, portanto, considerada uma melhor aproximação a uma futura teoria da informação pragmática. Para fins didáticos, a presente teoria da informação semântica pode ser identificada como uma teoria de informação pragmática para um receptor "ideal".

Parece desejável que alguém seja capaz de dizer não apenas que informação uma mensagem ou um experimento forneceu, mas também quanto. Daí, vamos distinguir entre informação (ou conteúdo) e quantidade de informação.

Devemos lidar com esses conceitos apenas na medida em que se aplicam a sentenças ou proposições, onde "sentenças" é abreviação para "sentenças

[*] Relatório técnico n. 247 do Laboratório de Pesquisa Eletrônica do MIT, publicado em 27 de outubro de 1952. A pesquisa relatada neste documento foi possível através do apoio estendido ao Massachusetts Institute of Technology, Laboratório de Pesquisa de Eletrônica, em conjunto com o Corpo de Sinalização do Exército, Departamento da Marinha (Escritório de Pesquisa Naval) e a Força Aérea (*Air Materiel Command*), sob o Corpo de Sinalização Contrato No. DA36-039 sc-100, Projeto n °. 8-102B-0; Departamento do Projeto do Exército nº 3-99-10-022; e em parte por uma concessão da Fundação Rockefeller.

ANALYTICA, Rio de Janeiro, vol 23 nº 2, 2019, p. 134-198
http://dx.doi.org/10.35920/1414-3004.2019v23n2p134-198

declarativas" ou "enunciados", e proposições são as entidades não linguísticas expressas por sentenças. A teoria que iremos desenvolver pressupõe um certo sistemas linguístico e os conceitos básicos desta teoria serão aplicados às sentenças desse sistema. Esses conceitos, então, serão conceitos semânticos, intimamente conectados com certos conceitos de lógica indutiva, como mostraremos abaixo. Uma vez que a lógica indutiva foi tratada em profundidade por um de nós[1], faremos uso extensivo dos resultados alcançados. Definições relevantes e teoremas serão, no entanto, repetidos a ponto de tornar o tratamento atual quase completamente independente.

A restrição do alcance de aplicação dos conceitos a serem explicados para frases (ou proposições) provavelmente não é grave, já que outras aplicações parecem ser redutíveis a esta. Em vez de lidar com as informações transportadas por letras, ondas sonoras, e assim por diante, podemos falar sobre a informação transportada pela sentença, 'A sequência de letras (ou ondas sonoras, etc.)... foi transmitida'. A situação é semelhante ao que prevalece no que diz respeito ao conceito de verdade, que é usado de forma sistemática como se aplicando não apenas a sentenças ou proposições, mas também a muitas outras entidades como conceitos e ideias. Nestas, também, estes últimos usos parecem ser redutíveis aos primeiros.

Em recentes apresentações autênticas da chamada Teoria Matemática da Comunicação, ou Teoria da (Transmissão de) Informação, grande cuidado foi tomado para ressaltar que esta teoria não está interessada nos aspectos semânticos da comunicação.

> As duas citações seguintes podem ser consideradas representativas. Claude E. Shannon declara em *The Mathematical Theory of Communication* (A Teoria Matemática da Comunicação da qual ele é co-autor com Warren Weaver), Univ. de Illinois Press, Urbana, 1949, p. 3: "Esses aspectos semânticos da comunicação são irrelevantes para o problema da engenharia." E. Colin Cherry em" *A History of the Theory of Information* ", Proc. Inst. Elec. Eng. 98, 383, 1951, diz:" É importante para enfatizar, no início, que não estamos preocupados com o significado ou a verdade das mensagens; a semântica está fora do escopo da teoria da informação matemática."

Entretanto, tem sido frequentemente notado que esse ascetismo nem sempre é respeitado na prática e que às vezes conclusões semanticamente importantes

[1] CARNAP. *Logical Foundations of Probability* (1950) e *The Continuum of Inductive Methods* (1952a). Estes trabalhos serão referidos a seguir como [Prob.] e [Cont.], respectivamente.

são tiradas de pressupostos oficialmente sem semântica. Além disso, parece que, pelo menos alguns dos proponentes da teoria da comunicação tentaram estabelecer (ou restabelecer) as conexões semânticas que foram deliberadamente rompidas por outros.

Em 1948, Donald MacKay concebeu uma teoria da informação que deveria ser suficientemente ampla para cobrir tanto a teoria da comunicação quanto a teoria da informação científica, a última lidando com a formação de representações ou o seu rearranjo na representação do espaço do observador, o primeiro tratando da replicação de representações na mente do receptor, que já estavam presentes na mente do remetente de uma mensagem. (Cf. "Aspectos Quânticos da Informação Científica", *Phil. Mag*. 41, 289, 1950, e "The Nomenclature of Information Theory", preparada para o Simpósio sobre a Teoria da Informação realizada em Londres em setembro de 1950, e impressa em formato revisado como Anexo I de uma palestra "*In Search of Basic Symbols*", dada antes da Oitava Conferência Cibernética, realizada em Nova York em março de 1951, e publicada no site *Transactions* desta conferência, pp. 181-235.)

Jean Ville, em palestra antes do 18º Congresso Internacional de Filosofia da Ciência em Paris, 1949 (publicado em *Actualites Scientifiques et Industrielles*, No. 1145, pp. 101-114, Paris, 1951), também trata a informação como um conceito basicamente semântico e desenvolve funções e teoremas que estão em estreita correspondência com algumas das funções e teoremas com os quais lidamos neste relatório. Uma avaliação mais completa desta e outras contribuições para os fundamentos da teoria da informação, bem como a comparação entre a teoria aqui apresentada e a teoria da comunicação será realizada em outro lugar.

Nossa teoria reside explicitamente e totalmente dentro da semântica.

> Não se trata, no entanto, com o que foi chamado por Weaver em sua contribuição ao livro acima mencionado "o problema semântico" da comunicação, que, como definido por ele, está "preocupado com a identidade, ou satisfatoriamente aproximação, na interpretação do significado pelo receptor, quando comparado com o significado pretendido do remetente. "Preferimos considerar uma investigação em que o remetente e o destinatário estão explicitamente envolvidos como pertencentes à pragmática.

Vamos falar sobre a informação transportada por uma sentença, tanto por si mesma como em relação a alguma outra sentença ou conjunto de sentenças, mas não sobre a informação que o remetente pretende enviar pela transmissão de uma certa mensagem nem sobre a informação obtida por um receptor a partir desta mensagem. Uma explicação desses usos é de primordial importância, mas é nossa convicção que a melhor abordagem para esta explicação seja através de uma análise do conceito de informação semântica que, além de ser uma aproximação, por abstração ao mais puro conceito de informação pragmática, pode bem ter seus próprios valores independentes.

Antecipando resultados posteriores, haverá, sob todas as explicações por nós previstas, que a quantidade de informações transportadas pela frase ' 17 X 19 = 323 ' é zero e que a quantidade de informação de ' As três medianas dos lados de um triângulo plano se cruzam em um ponto ', em relação a algum conjunto de frases servindo como um conjunto completo de axiomas para a geometria euclidiana, é igualmente zero. Isto, no entanto, não é de forma a ser compreendido como implicando que não haja um bom sentido de "quantidade de informação", em que a quantidade de informação dessas frases não será zero em absoluto, e para algumas pessoas, pode até ser bastante elevada. Para evitar ambiguidades, usaremos o adjetivo ' semântica ' para diferenciar tanto os sentidos pré-sistemáticos de informação" em que estamos interessados no momento e sua *explicata* sistemática de outros sentidos (como "quantidade de informação psicológica para a pessoa P") e sua *explicata*. Este adjetivo vai, no entanto, ser descartado nos casos em que as ambiguidades não sejam passiveis de surgir.

A comparação a seguir pode ser de valor para apontar um dos serviços que um esclarecimento do conceito semântico de informação deve render a uma futura teoria da informação pragmática. A teoria dos chamados gases ideais é de grande importância na física, apesar do fato de que nenhum gás real é ideal e que muitos gases estão muito longe de ser ideais. A informação semântica transportada por uma sentença em relação a uma certa classe de sentenças pode bem ser considerada como a informação pragmática "ideal" a qual esta sentença levaria para um receptor "ideal", cujo único conhecimento empírico é formulado exatamente nesta classe de sentenças. Por um receptor "ideal", entendemos, para os fins desta ilustração, um receptor com uma memória perfeita que "conhece" tudo de lógica e matemática, e juntamente com qualquer classe de sentenças empíricas, todas as suas consequências lógicas. A interpretação da informação semântica com a ajuda de tal intelecto fictício sobre-humano deve ser tomada apenas como uma indicação informal. Nós não iremos nos referir a esta ficção na parte técnica deste artigo.

A nossa tarefa pode agora ser declarada muito mais especificamente. Pretendemos explicar o conceito pré-sistemático de informação, na medida em que é aplicado a sentenças ou proposições e na medida em que se abstrai das condições pragmáticas da sua utilização. Então iremos definir, com base neste conceito sistemático de informação semântica, várias *explicata* para o conceito (ou conceitos) pré-sistemático de quantidade de informação semântica e investigaremos a sua adequação e aplicabilidade.

§2. Explicações Gerais

Os sistemas linguísticos relativos aos quais a atual teoria da informação é desenvolvida são descritos como contendo um número finito de constantes individuais e predicados primitivos de um lugar. Os seguintes conceitos fundamentais sintáticos e semânticos são explicados: sentença atômica, sentença molecular, sentença básica, predicado molecular, L-verdadeira, L-falsa, factual, L-implica, L-equivalente, L-disjunta, L-exclusiva, Q -predicador, Q -propriedade, Q -sentença, descrição-de-estado, e alcance. Alguns termos e símbolos da teoria de classes (teoria de conjuntos) são introduzidos, principalmente complemento, soma e produto.

Os sistemas linguísticos relativos aos quais nossa teoria da informação será desenvolvida são muito simples, tão simples que os resultados a serem obtidos serão de valor restrito em relação a sistemas linguísticos complexos o suficiente para servir como possíveis linguagens da ciência. A restrição, no entanto, foi parcialmente imposta pelo fato de que a lógica indutiva - sobre a qual teremos que confiar fortemente - foi até agora desenvolvida em um grau suficientemente elaborado apenas para idiomas que não são muito mais ricos do que aqueles tratados aqui, ([Prob.] §§ 15, 16), e em parte por uma questão de simplicidade de apresentação. Espera-se que, apesar disso, os resultados sejam imediatamente aplicáveis a certas situações simples e serão sugestivas em relação às mais complexas.

Nossos sistemas linguísticos $\mathcal{L}_n^\pi$ contêm n constantes individuais diferentes que representam n indivíduos diferentes (coisas, eventos ou posições) e π predicados primitivos de um lugar que designam propriedades primitivas dos indivíduos. (n e π são números finitos; sob certas assunções, no entanto, é fácil estender os resultados obtidos aqui para sistemas com um número infinitamente enumerável de constantes individuais). Em uma sentença atômica, por exemplo, 'P*a*' ('o indivíduo *a* tem a propriedade P'), uma propriedade primitiva é asserida como aplicando-se a um individuo. Outras sentenças moleculares são formadas de sentenças atômicas com a ajuda dos seguintes cinco conectivos habituais:

$\sim$	não	negação
$\vee$	ou	disjunção
$\wedge$	e	conjunção
$\rightarrow$	se...então	implicação (material)
$\leftrightarrow$	se, e somente se (escrito sse)	equivalência (material)

Todas as sentenças atômicas e suas negações são chamadas sentenças básicas. Analogamente, outros predicados moleculares ou predicadores são formados a partir de predicados primitivos com a ajuda dos conectivos (tipograficamente) iguais (por exemplo, '$M \wedge \sim P$' estando para 'M e não P'). Uma frase que consiste em um predicador e uma constante individual é chamada de uma sentença completa deste predicador. Embora nossos sistemas não contenham variáveis individuais, quantificadores, ou um sinal de identidade, seu poder expressivo não é, portanto, essencialmente afetado. Frases como 'existem exatamente três indivíduos com a propriedade P' podem ainda ser feitas nesses sistemas, embora apenas sob a forma um pouco desajeitada de uma disjunção de conjunções de frases básicas. Daí frequências absolutas (números cardinais de classes ou propriedades) e frequências relativas poderem ser expressas nestes sistemas (mas não quantidades mensuráveis como comprimento e massa).

Qualquer sentença é L-verdadeira (logicamente verdadeira, analítica, por exemplo, '$Pa \vee \sim Pa$') ou L-falsa (logicamente falsa, autocontraditória, por exemplo, '$\mathrm{Pa} \wedge \sim \mathrm{Pa}$') ou factual (logicamente indeterminada, sintética, por exemplo, '$Pa \vee [M \wedge \sim \mathrm{N}]\mathrm{b}$'). Relações lógicas entre sentenças i e j podem ser definidas:

$$
\begin{array}{lll}
\mathrm{i\ L-implica\ j} & =_{\mathrm{DF}} & \mathrm{i} \rightarrow \mathrm{j\ é\ L-verdadeiro} \\
\mathrm{i\ é\ L-equivalente\ a\ j} & =_{\mathrm{DF}} & \mathrm{i} \leftrightarrow \mathrm{j\ é\ L-verdadeiro} \\
\mathrm{i\ é\ L-disjunto\ com\ j} & =_{\mathrm{DF}} & \mathrm{i} \vee \mathrm{j\ é\ L-verdadeiro} \\
\mathrm{i\ é\ L-exclusivo\ de\ j} & =_{\mathrm{DF}} & \mathrm{i} \wedge \mathrm{j\ é\ L-verdadeiro}
\end{array}
$$

Usaremos 't' como o nome de uma sentença L-verdadeira em particular, uma "tautologia", digamos, de '$Pa \vee \sim Pa$'.

Um Q -predicador é uma conjunção (de predicados) na qual cada predicado primitivo ocorre não negado ou negado (mas não ambos) e nenhum outro predicado ocorre.

A propriedade designada por um Q -predicador é chamada de Q -propriedade. Uma sentença completa de um Q-predicador é uma Q-sentença. Uma descrição-de-estado[2] é uma conjunção de n Q-sentenças, uma para cada indivíduo. Assim, uma descrição-de-estado descreve completamente um possível estado do universo do discurso em questão.[3] Para qualquer sentença j do sistema, a classe daquelas descrições-de-estado em que j contém, isto é, cada uma das quais L-implica j, é chamada o alcance de j. O intervalo de j é nulo se, e somente se, j for L-falsa; em qualquer outro caso, j é L-equivalente à disjunção das descrições-de-estado em seu alcance.

Os seguintes teoremas serão de uso posterior:

T2-1

a. O número de sentenças atômicas é $\beta = \pi\, n$
b. O número de Q –predicadores é $\kappa = 2^{\pi}$
c. O número de descrições-de-estado é $z = 2^{\beta} = 2^{\pi\, n} = (2^{\pi})^{n} = \kappa^{n}$

Na nossa metalinguagem, isto é, a língua em que falamos sobre os nossos sistemas de linguagem $\mathcal{L}_n^{\pi}$ (no nosso caso, uma certa sub-linguagem não especificada do português [inglês] comum enriquecida por alguns símbolos adicionais), usaremos alguns termos e símbolos usuais da teoria de classes (ou conjuntos). A classe de todas essas entidades (de um certo tipo) que não pertencem a uma determinada classe K será chamado o complemento classe de K e denotado por '-K'. A classe das entidades pertencentes a uma classe K ou a uma classe L (ou a ambos) será chamada a soma (classe-teórica) dessas classes e será denotada por '$K \cup L$'. A classe das entidades que pertencem a cada uma das classes K e L será chamada de produto (classe-teórico) dessas classes e denotada por '$K \cap L$'.

Aqueles leitores que podem não estar familiarizados com conceitos e termos lógicos abstratos vão se beneficiar com a ilustração a seguir, o que será utilizada ao logo dessa monografia. Seja um censo feito em uma pequena comunidade de apenas três habitantes, e deixando-se interessar apenas em saber se os habitantes contados são do sexo masculino ou não-masculino (feminino) e jovem ou não-jovem (idoso), respectivamente. Sejam os três indivíduos serem designados por

[2] N.T. a expressão original é *state-description*

[3] Isso, estritamente falando, é válido apenas se as propriedades primitivas forem logicamente independentes. Para uma discussão dos problemas envolvidos, veja CARNAP 1952b e a literatura ali mencionada.

'*a*', '*b*', '*c*' e as propriedades por ‘M’, ‘~M’ (ou ‘F’), ‘Y’, e ‘~Y’ (ou ‘O’)[4], respectivamente. O sistema linguístico, no qual o resultado do censo pode ser exaustivamente descrito é, portanto, um sistema $\mathcal{L}_3^2$, em nossa notação. ‘Ma’ é uma sentença atômica, ‘Ma ∧ Fb ∧ (Mc → 0c)’ outra sentença molecular, ‘F ∧ ~Y’ um Q-predicador, ‘[F ∧ ~Y]b’ uma Q-sentença, ‘[M ∧ Y]a ∧ [~M ∧ Y]b ∧ [~M ∧ ~Y]c’ uma descrição-de-estado [*state-description*]. Para posteriores referências, a lista de todas as 64 descrições-de-estado é dada na Tabela I, de forma abreviada. A linha 10 desta tabela, por exemplo, deve ser interpretada como abreviada para a descrição do estado **F**. Mais tarde (§4), no entanto, uma interpretação diferente da mesma tabela será dada.

	M, Y	M, O	F, Y	F, O
1.	a, b, c	-	-	-
2.	-	a, b, c	-	-
3.	-	-	a, b, c	-
4.	-	-	-	a, b, c
5.	a, b	c	-	-
6.	a, c	b	-	-
7.	b, c	a	-	-
8.	a, b	-	c	-
9.	a, c	-	b	-
10.	b, c	-	a	-
11.	a, b	-	-	c
12.	a, c	-	-	b
13.	b, c	-	-	a
14.	c	a, b	-	-
15.	b	a, c	-	-
16.	a	b, c	-	-
17.	-	a, b	c	-
18.	-	a, c	b	-
19.	-	b, c	a	-
20.	-	a, b	-	c
21.	-	a, c	-	b
22.	-	b, c	-	a
23.	c	-	a, b	-
24.	b	-	a, c	-
25.	a	-	b, c	-
26.	-	c	a, b	-
27.	-	b	a, c	-
28.	-	a	b, c	-
29.	-	-	a, b	c
30.	-	-	a, c	b
31.	-	-	b, c	a
32.	c	-	-	a, b

	M, Y	M, O	F, Y	F, O
33.	a	-	-	a, c
34.	b	-	-	b, c
35.	-	c	-	a, b
36.	-	b	-	a, c
37.	-	a	-	b, c
38.	-	-	c	-
39.	-	-	b	-
40.	-	-	a	-
41.	a	b	c	-
42.	a	c	b	-
43.	b	a	c	-
44.	b	C	a	-
45.	c	a	b	-
46.	c	b	a	-
47.	a	b	-	c
48.	a	c	-	b
49.	b	a	-	c
50.	b	c	-	a
51.	c	a	-	b
52.	c	b	-	a
53.	a	-	b	c
54.	a	-	c	b
55.	b	-	a	c
56.	b	-	c	a
57.	c	-	a	b
58.	c	-	b	a
59.	-	a	b	c
60.	-	a	c	b
61.	-	b	a	c
62.	-	b	c	a
63.	-	c	a	b
64.	-	c	b	a

[4] (N.T.) ‘M’ para *male*, ‘F’ para *female*, ‘Y’ para *young* e ‘O’ para *old*.

O leitor irá verificar facilmente que o alcance [*range*] da sentença 'Ma ∧ Ya ∧ Fb ∧ Yb' que pode ser reescrita vantajosamente na forma '[M ∧ Y]a ∧ [F ∧ Y]b', contém exatamente 4 descrições-de-estado, nomeadamente, 9, 25, 42 e 53. O alcance de 'Fa' contém 32 descrições-de-estado. O alcance de 'Ma ∨ Ya ∨ Fb ∨ Yb ∨ Fc ∨ 0c' contém 63 descrições-de-estado, ou seja, todas as descrições-de-estado, exceto 52. Um leitor com algum treinamento na lógica proposicional verá imediatamente que esta última sentença é L-equivalente a ~(Ma ∧ Ya ∧ Fb ∧ Yb ∧ Fc ∧ 0c), portanto, [equivalente] à negação da descrição-de-estado 52.

§3. O conceito pré-sistemático de informação semântica

> Um requisito de adequação para qualquer explicação proposta de informação semântica - In - é declarado: In (i) inclui In (j) se e somente se i L- implica j. A partir deste requisito vários teoremas são derivados. Além da informação absoluta transportada por uma sentença, a informação transportada por uma sentença j em excesso ao que é transportado por alguma outra sentença, muitas vezes é importante. Este conceito de informação relativa é definido por: In (j / i) = In ($i \wedge j$) - In (i). Um dos teoremas é: se t é qualquer sentença L-verdadeira, In (j / t) = In (j). Dois conceitos que cumprem o requisito, mas diferem em algum aspecto são investigados, mas nenhum deles aceitos como *explicatum* para In.

Para dispersar, pelo menos parcialmente, a dúvida que envolve as inevitavelmente vagas discussões sobre a adequação da explicação a ser oferecida posteriormente para o conceito de informação, deixe-nos indicar um requisito que servirá como uma condição necessária para essa adequação.

Sempre que i L-implica j, i afirma tudo o que é afirmado por j e possivelmente mais. Em outras palavras, a informação transportada por i inclui as informações transportadas por j como uma parte (talvez imprópria). Usando 'In (...)' como uma abreviatura para o conceito pré-sistemático 'as informações transportadas por. . . ', agora podemos declarar o requisito do seguinte modo:

R3-1 In(i) inclui In(j) sse i L-implica j.

Por esta exigência nos comprometemos a tratar a informação como um conjunto ou classe de alguma coisa. Isso está de acordo com as formas comuns de expressão, como por exemplo, "As informações fornecidas por esta afirmação são mais inclusivas do que (ou é idêntico ou sobrepõe-se) as fornecidas pela outra afirmação."

Vamos agora declarar alguns teoremas que valem para 'In' e, portanto, também para aquele conceito que ofereceremos, na seção seguinte, como o *explicatum* de 'In'. Estes teoremas seguem os de R*i*[5] * e os teoremas conhecidos da teoria de classes.

T3-1. In(i) = In(j) sse i é L-equivalente a j.

Se uma classe K de classes contém uma classe que está incluída em todos os membros de K, esta classe será chamada "a classe mínima (de K)". Se K contiver uma classe que inclui todos os membros de K, esta classe será chamada "a classe máxima (de K)". A classe mínima e a classe máxima podem coincidir com a classe nula e a classe universal (do tipo correspondente), respectivamente, mas não precisam sê-lo.

Como uma sentença L-verdadeira é L-implicada por cada sentença, e uma sentença L -falsa L- implica cada sentença, temos:

T3-2. In(i) = a classe mínima de K (onde K é agora a classe das In -classes de sentenças) sse i é L- verdadeira.

T3-3. In(i) = a classe máxima de K sse i é L-falsa.

Talvez, a princípio, pareça estranho que uma sentença autocontraditória, portanto uma que nenhum receptor ideal aceitaria, seja considerada como a informação mais inclusiva. Deve-se, no entanto, enfatizar que a informação semântica aqui não significa implicação da verdade. Uma frase falsa que, por acaso, diz muito é altamente informativa em nosso sentido. Se a informação que carrega é verdadeira ou falsa, cientificamente valiosa ou não, e assim por diante, não nos diz respeito. Uma frase autocontraditória afirma demais; é muito informativa para ser verdade.

T3-4. In (i) inclui propriamente In (j) sse i L-implica j mas j não implica L-i.

T3-5. In (i) inclui propriamente a classe mínima e está devidamente incluída na classe máxima sse i é factual.

[5] Por razões de brevidade, teoremas, definições ou requisitos, quando encontrados na mesma seção em que são declarados pela primeira vez, serão referidos apenas pela letra 'T', 'D' ou 'R' e pelos seus segundos números. Aqui, por exemplo, nós temos 'R1' em vez do maior 'R3-1'.

T3-6. In(i) inclui In($i \vee j$) e está incluída em In($i \wedge \ j$).

Quando usamos o termo "informação" na linguagem comum, muitas vezes nos referimos à informação carregada por uma sentença absolutamente, por assim dizer. Pelo menos com tanta frequência, no entanto, tenciona referir-se às informações transportadas por uma sentença que exceda a alguma outra sentença (ou classe de sentenças). Se não for indicado de outra forma ou implicitamente entendido através do contexto, esta outra sentença será muitas vezes aquela na qual o conhecimento total disponível ao receptor da informação, antes de receber a nova informação, é indicado. Em contradição com o conceito de informação absoluta tratado até o momento, vamos agora definir, ainda no nível pré-sistemático, o conceito de informação relativa (ou adicional ou excessiva) de j em relação a i como a diferença classe-teórica de In $In(i$ i∧j $j)$ e In (i), isto é, o produto classe-teórico de In $In(i$ i∧j $j)$ com o complemento de In(i); em símbolos:

D3-1. In $(\mathrm{j}/\mathrm{i}) = \mathrm{In}(\mathrm{j}/\mathrm{i})\ - \mathrm{In}(\mathrm{j})\ \big(= \mathrm{In}(\mathrm{i} \wedge \mathrm{j}) \cap -\mathrm{In}(\mathrm{i})\big)$

$In\ (j/i)$ é novamente uma classe. Seus membros pertencem ao mesmo tipo que os membros do In(i). Os seguintes teoremas seguem imediatamente de D1, R1 e os teoremas anteriores.

> Provas formais completas serão dadas somente quando uma indicação dos teoremas, definições e requisitos a partir dos quais um teorema siga não permita a maioria dos leitores compreender a prova por inspeção. Em casos muito simples (como em os primeiros 8 teoremas), estas dicas serão omitidas.

T3-7. $In(j/i)$ inclui a classe mínima e está incluída na classe máxima.

T3-8. Se i é L-equivalente a j, então In(k/i) = In(k/j) e In(i/ℓ) = In(j/ℓ)

T3-9. Se i-L implica j, então $In(j/i)$ = a classe mínima.

Prova: Neste caso, $j \wedge i$ é L-equivalente a i. O teorema segue de T1 e D1

T3-10. Se j é L- verdadeira, então $In(j/i)$ = a classe mínima.

T3-11. Em (j/i) inclui propriamente a classe mínima sse i não L-implica j.

Prova: Nesse caso, $In(j \wedge i)$ inclui corretamente In (j).

T3-12.

a. Se i é uma sentença L- verdadeira, $In(j/i)$ = In(j).

Prova: Neste caso, j. i é L-equivalente a j e In(i) = a classe mínima.

b. $In(j \wedge i)$ = In(j).

Assim, a informação relativa de *j* em relação a *t* é igual a informação absoluta de *j*. Portanto, seria possível começar com a informação relativa como primitiva e definir a informação absoluta como o valor da informação relativa com respeito a *t*. No entanto, parece mais conveniente começar com o conceito simples de informação absoluta, porque esta tem somente um argumento e a informação relativa pode ser definida com base no mesmo. Este é o procedimento que escolhemos aqui.

Até agora, nos comprometemos a tratar a informação transportada por uma sentença como uma classe de algo e declaramos um requisito o qual todo *explicatum* adequado terá que cumprir. Isto, claro, deixa muitas possibilidades abertas. Com respeito a informação realizada por uma sentença L-verdadeira, somos capazes de afirmar apenas que é um mínimo e está contida nas informações transportadas por qualquer sentença. Talvez pareça plausível exigir, além disso, que a informação transportada por uma sentença L-verdadeira deva ser vazia; portanto, a classe nula do tipo apropriado. Mas essa sensação é devida ao fato de que nem sempre distinguimos cuidadosamente entre informação e quantidade de informação. O que realmente temos em mente é que a quantidade de informação semântica carregada por uma sentença de L-verdadeira deva ser zero. Mas isso pode ser alcançado por um *explicatum* adequado, mesmo que as informações transportadas por essa frase não sejam a classe nula.

Por outro lado, também não existe nenhuma boa razão até agora de porque a informação de uma sentença L-verdadeira não deva ser a classe nula. O melhor procedimento é, portanto, deixar esta decisão em aberto.

Há realmente duas *explicata* plausíveis para In(*i*), que diferem exatamente neste ponto: de acordo com uma, a informação carregada por uma sentença de L-verdadeira será a classe nula; de acordo com a outra, não. Vamos denotar o primeiro conceito por 'Inf1' e o segundo por 'Inf2' ", com o 1 e o 2 subscritos, tal como em D3-2 e D3-3, respectivamente:

D3-2. $Inf_1(i) =_{DF}$ a classe de todas as sentenças $(em\ \mathcal{L})$ que são L-implicadas por i e não L-verdadeiras.

D3-3. $Inf_2(i) =_{DF}$ a classe de todas as sentenças ($em\ \mathcal{L}$) que são L-implicadas por i.

Não vamos nos debruçar sobre uma comparação elaborada dos méritos e falhas relativas destas duas definições: primeiro, porque essa comparação já foi realizada por um de nós em um contexto[6]* intimamente relacionado, em segundo lugar, porque não devemos adotar nenhuma destas definições para o trabalho futuro, mas uma terceira a ser explicada na seção seguinte.

§4. Elementos-de-conteúdo e conteúdo

> Um elemento-de-conteúdo [*content-element*] é definido como a negação de uma descrição-de-estado e o conteúdo de *i*-Cont (*i*)- como a classe dos elementos-de-conteúdo L-implicada por *i*. Cont é tomado como o *explicatum* para In. Cont (j/i) é definido e vários teoremas derivados.

No §2, definimos a ordem de uma sentença *i*, R(*i*), como a classe de todas as descrições-de-estado Z em que *i* possui ou que, em outras palavras, L-implica *i*. A sentença *i* diz que o estado do universo (tratado em $\mathcal{L}$) é um dos estados possíveis que são descritos por Z em R(*i*). Alternativamente formulado, *i* diz que o universo não está em nenhum dos estados que são descritos por Z em $V_Z - R(i)$, onde V_Z é a classe de todo Z. Assim como *i* é L- implicada por cada Z em R(i), então L-implica a negação de cada Z em $V_Z - R(i)$. Chamamos essas negações os elementos-de-conteúdo E de *i* e sua classe o conteúdo de *i*, em símbolos Cont (*i*). Em geral, chamamos as negações do Z, em um determinado sistema $\mathcal{L}$, o E deste sistema. (vide [Prob.] § 73.)

Em nosso $\mathcal{L}_3^2$ há, naturalmente, 64 elementos-de-conteúdo [*content-elements*], ou seja, as negações de suas 64 descrições-de-estado. Esses elementos-de-conteúdo aparecem na tabela I, quando interpretados de uma maneira diferente daquela dada antes. Agora podemos ler a linha 10, por exemplo, como '[M ∨ Y]b ∨ [M ∨ Y]c ∨ [F ∨ Y]a', um elemento de conteúdo que é L-equivalente à negação da descrição do estado 37, como o leitor irá verificar por si mesmo.

---

[6] Vide CARNAP (1942), §23, and [Prob.] , p. 406.

O conteúdo da frase ‘Ma ∧ Ya ∧ Fb ∧ Yb’ contém 60 elementos-de-conteúdo, nomeadamente, as negações de todas as descrições-de-estado, exceto 9, 25, 42 e 53.

O seguinte teorema, T4-1, pode ser deduzido dos teoremas sobre descrições-de-estado e L-conceitos em [Prob.] § § 18A, 18D, 19, 20, 21B.

T4-1. Para cada Ei as seguintes considerações:

a. E_i é factual ([Prob.] T20-5b, T20-6).

b. Se E_j é distinto de E_i, então E_i e E_j são L-disjuntos.

Prova: $\sim E_i \wedge \sim E_j$ é L-falsa ([Prob.] T20-8a). Portanto, a negação desta conjunção é L-verdadeira. Mas essa negação é L-equivalente a $E_i \vee E_j$.

c. A conjunção de todos E_i é L-falsa.

Prova: deixe d ser a disjunção das negações do E_i, consequentemente L-equivalente à disjunção de todos Z. Portanto, d é L- verdadeira ([Prob.] T21-8b); consequentemente ~d é L-falsa. Mas ~d é L-equivalente à conjunção de todas E_i.

d. Se E_i L-implica j, então j é ou L- verdadeira ou L- equivalente a E_i; em outras palavras, E_i é uma sentença factual mais fraca.

Assim como uma descrição-de-estado diz o máximo que pode ser dito no dado universo de um discurso, escasso de contradição, então um elemento-de-conteúdo, diz o mínimo, além de uma tautologia. ' *a* é do sexo masculino e jovem, *b* é do sexo feminino e jovem, e *c* é do sexo feminino e velho ' é uma sentença factual mais forte no censo; sua negação ' *a* é feminino ou velho (ou ambos), ou *b* é masculino ou velho, ou *c* é masculino ou jovem (onde ' ou ' é sempre entendido em seu sentido não-exclusivo) um mais fraco.

T4-2.

a. Cont (i) = a classe nula de E, $\wedge_E$ sse i é L- verdadeira.

b. Cont (i) = a classe de todos os E, $\vee_E$ sse i é L-falsa.

c. Cont (i) = nem $\wedge_E$ nem $\vee_E$ sse i é factual.

d. Cont (i) inclui Cont (j) sse i L-implica j.

e. Cont (i) = Cont (j) sse i é L-equivalente a j.

f. Cont (i) e Cont (j) são exclusivos (quer dizer, não têm membros em comum) sse i e j são L-disjuntas ([Prob.] T20-lc, d).

Os conteúdos de ‘Ma’e de ‘Fa ∨ Mb’ são exclusivos, já que ‘Ma ∨ Fa ∨ Mb’ é L-verdadeira.

O leitor pode verificar na Tabela I, em sua segunda interpretação, que esses conteúdos na verdade não têm membros em comum.

T4-3.

a. Cont(~i) = −Cont (i) (abreviação de '−Cont (i)', [Prob.] T18-le).

b. $\mathrm{Cont}(\mathrm{i} \vee \mathrm{j}) = \mathrm{Cont}(\mathrm{i}) \cap \mathrm{Cont}(\mathrm{j})$

Prova: Seja $\overline{R} = (\dots)$ a classe das negações dos membros de $R = (\dots)$. Então
$$\mathrm{Cont}(\mathrm{i} \vee \mathrm{j}) = \overline{\mathrm{R}}\big({\sim}(\mathrm{i} \vee \mathrm{j})\big) = \overline{\mathrm{R}}({\sim}\mathrm{i} \wedge {\sim}\mathrm{j}) = \overline{\mathrm{R}}({\sim}\mathrm{i}) \cap \overline{\mathrm{R}}({\sim}\mathrm{j}) = \mathrm{Cont}(\mathrm{i}) \cap \mathrm{Cont}(\mathrm{j})$$

c. $\mathrm{Cont}(\mathrm{i} \wedge \mathrm{j}) = \mathrm{Cont}(\mathrm{i}) \cup \mathrm{Cont}(\mathrm{j})$.

Prova: $\mathrm{Cont}(\mathrm{i} \wedge \mathrm{j}) = -\mathrm{Cont}\big({\sim}(\mathrm{i} \wedge \mathrm{j})\big) = -\,\mathrm{Cont}\,(\,{\sim}\mathrm{i} \vee {\sim}\mathrm{j}\,) \;=\; -(\mathrm{Cont}({\sim}\mathrm{i})) \cap \mathrm{Cont}({\sim}\mathrm{j})) = -(-\mathrm{Cont}(\mathrm{i}) \cap -\mathrm{Cont}(\mathrm{j})) = \mathrm{Cont}(\mathrm{i}) \cup \mathrm{Cont}(\mathrm{j})$

Para verificar T3*b*, *c* tome, por exemplo, *i* como ‘Ma ∨ Fb ∨ [M ∨ Y]c’e *j* como ‘Fa ∨ Mb’.

O T2*d* mostra que Cont cumpre o requisito R3-1. Nós decidimos tomar Cont como nosso *explicatum* para In. A explicação da informação transportada por uma sentença *j*, como a classe das negações de todos aqueles Z que são excluídos por *j*, é intuitivamente plausível e de acordo com o antigo princípio filosófico, "*omnis determinatio est negatio*”. Nosso principal motivo, no entanto, para dar preferência sobre as duas *explicata* mencionadas na seção anterior, Inf_1 e Inf_2, reside no fato de que uma explicação de quantidade de informações se tornará bastante simples se baseada em Cont, de acordo com o quarto requisito para uma boa explicação declarada em [Prob.], p. 7.

Notemos que, de acordo com a T2*a*, Cont compartilha com Inf_1 que seus valores para uma sentença L-verdadeira como argumento é a classe nula.

Agora temos que definir o conteúdo relativo de *j* em relação a *i*. O que tem que ser feito, lógico, é simplesmente substituir 'In' em D3-1 por 'Cont'.

D4-1 $\mathrm{Cont}(\mathrm{j}/\mathrm{i}) =_{\mathrm{DF}} \mathrm{Cont}(\mathrm{i} \wedge \mathrm{j}) - \mathrm{Cont}(\mathrm{i})$

Claramente, T3-12*a*, *b* se aplica a Cont se 'In' for substituído neles por 'Cont'. Deixe-nos afirmar explicitamente apenas o correlato de T3-12*b*, isto é,

T4-4. Cont(j/t) = Cont(j).

As observações que se seguem ao T3-12*b* são também *mutatis mutandis* para conteúdo absoluto e relativo.

§5. O conceito pré-sistemático de quantidade de informação

> Requisitos de adequação para a explicação da quantidade de informação semântica - in - são declarados, e teoremas para in derivados. Nenhum requisito formal de aditividade é aceito, uma vez que as condições sob as quais a aditividade deve se manter não podem ser dadas sem ambiguidade, até o momento. In (j / i), a quantidade de informação de j em relação a i, é definida e teoremas derivados.

Nossa próxima tarefa é encontrar um *explicatum*, ou talvez várias *explicata*, para o conceito pré-sistemático de quantidade de informação. Isto será novamente precedido pela declaração de alguns requisitos, cujo cumprimento será condição necessária para a adequação da *explicata* a ser proposta.

Nós usaremos "in" como símbolo do conceito pré-sistemático de quantidade de informação e distinguir entre a quantidade absoluta de informação de uma frase i, em (i), e a quantidade relativa de informação da sentença j em relação a i, em (j/i). O montante relativo é claramente definível com base no valor absoluto:

D5-1. $in(j/i) =_{DF} in(i \wedge j) - in(i)$

(onde o sinal '-' é desta vez o símbolo da diferença numérica e não, como em D3-1 ou D4-1, para diferença de classe). Portanto, basta indicar apenas os requisitos em relação ao valor absoluto. Parece plausível exigir que a quantidade de informação de i não deva ser menos que a quantia de informação de j, se o conteúdo de i inclui o conteúdo de j; que a quantidade de informação de uma sentença L-verdadeira deva ser zero; e, para sistemas finitos, que a quantidade de informação de uma sentença deva ser maior que zero. (A qualificação 'para sistemas finitos' pode, talvez, parecer supérflua. Pode, no entanto, ser mostrado que, no que diz respeito à *explicata* prevista por nós, este requisito não seria cumprido em um sistema infinito.) Mais formalmente:

R5-1. in(i) $\geq$ in(j) se (mas não somente se) Cont(i) inclui Cont(j).

R5-2. in(j) = 0 se Cont(j) = Λ_E.

R5-3. in(j) $>$ 0 se Cont (j) incluir corretamente Λ_E.

Em vez de R3, também poderíamos ter exigido a seguinte condição um pouco mais forte da qual R3 segue imediatamente:

R5-4. $in(i) > in(j)$ se Cont(i) incluir propriamente Cont(j).

Poderíamos também ter declarado esses requisitos diretamente em termos de "L-implica" e 'L- verdadeira', sem recurso a Cont. Para o benefício daqueles que, por alguma razão, não estão satisfeitos com a nossa explicação de 'informação' e que, portanto, tentam explicar "quantidade de informação" com base em alguma outra explicação para "informação" ou talvez até mesmo sem referência a qualquer tal *explicatum* (um perfeitamente razoável objetivo alcançável), a seguinte versão é dada:

R5-1 *. $in(i) \geq in(j)$ se (mas não somente se) i L-implica j.

R5-2 *. in(j) = 0 se j é L-verdadeira.

R5-3 *. in(j) > 0 se j não for L- verdadeira.

Os teoremas a seguir se seguem de R1 a R3 e das propriedades previamente declaradas de Cont.

T5-1 Se Cont(i) = Cont(j), então in(i) = in(j).

T5-2. Se i for L-falsa, então in(i) tem o valor-in máximo.

Prova: Uma sentença L-falsa L -implica cada sentença.

T5-3. 0 < in(i) < o valor máximo de in sse i é factual.

T5-4. $in(i \vee j) \leq in(i) \leq in(i \wedge j)$

Os requisitos R1 a R3 são claramente bastante fracos e pode-se procurar outros requisitos. Um que se recomenda imediatamente seria o da aditividade, isto é, $ter\ in(i \wedge j) = in(i) + in(j)$, se i e j forem independentes um do outro em um certo sentido. No entanto, não faremos deste um dos nossos requisitos formais, porque o sentido da independência envolvida não está claro neste momento. Vamos encontrar mais tarde que cada uma das nossas *explicata* é realmente aditiva, mas nem todas elas no mesmo sentido, porque as condições de independência não são as mesmas nos vários casos.

A aditividade é válida apenas sob certas condições, quaisquer que sejam essas condições em termos exatos. É claro que, em geral, **F**. É claro que haverá casos em que in($i.\ j$) < in(i) + in (j). Este será o caso por exemplo, sempre que i L-implica j e j não é L-verdadeira, porque nestas circunstâncias i é L- equivalente a $i.\ j$, de modo que in ($i.j$) = in (i), enquanto in (j)> 0 e, portanto, in(i) < in(i) + in(j). Até agora, podemos indicar apenas um limite inferior para in ($i.j$), especificamente:

T5-5. $in(i \wedge j) \geq \max[in(i), in(j)]$

Existe um limite superior geral para $in(i \wedge j)$ que não seja trivial? Nenhum teorema para este efeito pode ser deduzido dos requisitos desta seção. Eles não excluem, por exemplo, a possibilidade de que algumas vezes $in\,(i \wedge j) > in\,(i) + in\,(j)$. Essa possibilidade pode talvez parecer tão implausível que alguém a excluiria pelo requisito adicional explícito $n\,(i \wedge j) \leq in\,(i) + in\,(j)$. No entanto, parece melhor não exigir isso. Veremos mais tarde que o segundo da nossa *explicata* (inf) viola esta condição, e devemos então tornar essa violação plausível. Se alguém insiste que o requisito que acabou de ser posto tem que ser cumprido, então ele pode aceitar somente o primeiro de nossa *explicata* (cont). Por este conceito a exigência é realmente cumprida (T6-4m).

Os seguintes teoremas correspondem a T3-7 até T3-12.

T5-7. O valor-in máximo $\geq in(j/i) \geq 0$.

T5-8. Se i é L-equivalente a j, então $in(k/i) = in(k/j)\ e\ in(i/\ell) = in(j/\ell)$.

T5-9. Se i-implica j, então $in(j/i) = 0$.

T5-10. Se j é L-verdadeira, então $in(j/i) = 0$.

T5-11. in(j / i)> 0 sse i não L- implica j.

T5-12.

a. Se i é uma sentença L-verdadeira, in(j/i) = in(j).

b. in(j/t) = in(j).

§ 6. *O primeiro explicatum: Medida de conteúdo (cont)*

Uma maneira de cumprir os requisitos declarados na seção anterior é descrita. Consiste, essencialmente, em definir uma função-medida sobre os elementos-de-conteúdo, cumprindo certas condições e, em seguida, tomando como medida do conteúdo de uma sentença a soma das medidas atribuídas aos elementos de seu conteúdo. Como as funções de medida sobre descrições-de-estado - funções m - já

> foram tratadas por um dos autores antes, uma maneira mais curta de introduzir medidas de conteúdo - cont - é escolhida, simplesmente equacionando cont(i) com $m_P(\sim i)$ onde m_P significa função-m própria, ou seja, função-m que preenche certas condições. Muitos teoremas para cont (i) são derivados, entre eles, os teoremas para as medidas de conteúdo de frases básicas, para disjunções e conjunções de tais, para sentenças Q, e para sentenças em forma normal disjuntiva e conjuntiva. $cont(j/i)$ é definido, e entre outros, o importante teorema $cont(j/i) = cont(i \rightarrow j)$ é derivado.

Poderíamos ter definido um *explicatum* adequado para a quantidade de informação transportada por uma frase com a ajuda de funções de medida que vão além do conteúdo das sentenças de $\boldsymbol{\mathcal{L}}$ e cumprindo as condições estabelecidas na seção anterior. No entanto, como existem relações estreitas entre os conteúdos e os intervalos (§4), utilizaremos o fato de que as definições de várias funções-medida nos intervalos já foram tratadas com profundidade em [Prob.] e definir as funções nas quais estamos agora interessados simplesmente com base nessas funções-medida.

Parece proveitoso começar com uma espécie de função-medida que abranja as afirmações e outras sentenças que não foram discutidas explicitamente em [Prob.] ou [Cont.], nomeadamente, com funções m próprias, a serem denotadas por 'm_P'.

Definimos:

D6-1 m é uma função-m própria $(em\ \mathcal{L}) =_{DF}$ m preenche as nove condições seguintes:

a. Para cada $Z_i, m(Z_i) > 0$.

b. A soma dos valores-m de todos os Z = 1.

c. Para qualquer sentença L-falsa j, m (j) = 0.

d. Para qualquer sentença não-L-falsa j, m (j) = a soma dos valores-m para o Z em R(j).

e. Se Z_j é formado a partir de Z_i, substituindo as constantes individuais de Z_i pelas correlacionadas para eles por qualquer permutação das constantes individuais, então $m(Z_j) = m(Z_i)$. (Menos estritamente, porém mais sugestivamente: todos os indivíduos são tratados em um par.)

f. Se Z_j é formado a partir de Z_i, substituindo os predicados primitivos de Z_i pelos correlacionados para eles por qualquer permutação dos predicados primitivos, então $m(Z_j) = m(Z_i)$. (ou seja, todas as propriedades primitivas são tratadas em um par).

g. Se Z_j é formado a partir de Z_i, substituindo qualquer um dos predicados primitivos de por suas negações (omitindo sinais de dupla negação), então $m(Z_j) = m(Z_i)$ (ou seja, cada propriedade primitiva é tratada em um par com o seu complemento).

As três últimas condições poderiam ter sido declaradas de forma um pouco mais fraca, mas nenhuma tentativa foi feita para reduzir a redundância, sacrificando a clareza psicológica.

h. Se i e j não tiverem predicados primitivos em comum, então $m\ (i \wedge j)\ = m\ (i) \times m\ (j)$.

i. $m(i)$ não é influenciado pelo número de indivíduos de $\boldsymbol{\mathcal{L}}$ não mencionados em i. (Esta condição será usada apenas na derivação da fórmula (6) no §10.)

Uma função m cumprindo as condições (*a*) até (*d*) é chamada regular ([Prob.] P. 295). Se preencher, além disso, a condição (*e*), é chamada simétrica ([Prob.] P. 485). Todos os teoremas que guardam funções m regulares sustentam a fortiori qualquer função m adequada.

Acredita-se que m_P seja um *explicatum* adequado para um dos sentidos em que "probabilidade" é usada, ou seja, o que poderia ser denominado "probabilidade lógica absoluta", isto é, probabilidade lógica sem evidência (ou evidência tautológica ou evidência irrelevante).

Similarmente, c_P, a ser definido em D7-3, acredita-se ser um *explicatum* de probabilidade lógica relativa.

Quaisquer duas sentenças (não apenas descrições-de-estado) que estejam na relação declarada em Dle são chamadas isomórficas

O seguinte teorema vale para todas as funções m regulares ([Prob.] §§ 55A, 57A), daí também para todas as funções m próprias:

T6-1

a. $0 \leq m(i) \leq i1$

b. $m(i)\ =\ 1$ sse i é L-verdadeira.

c. $m\ (i)\ =\ 0$ sse i é L-falsa.

d. $0 < m(i) < 1$ sse é factual.

e. Se i L-implica j, então $m(i) \leq m(j)$.

f. Se i é L-equivalente a j, então $m(i) = m(j)$.

g. $m(i \wedge j) \leq m(i) \leq m(i \vee j)$.

h. $m(i \vee j) = m(i) + m(j) - m(i \wedge j)$.

i. $m(i \vee j) = m(i) + m(j)$ sse i. j é L-falsa (isto é, sse i e j são L-exclusivos).

j. $m(i \wedge j) = m(i) + m(j) - m(i \vee j)$.

k. $m(i \wedge j) = m(i) + m(j) - 1$ sse $i \vee j$ é L-verdadeira (isto é, sse i e j são L-disjuntas).

1. $m(\sim i) = 1 - m(i)$.

m. $m(i \wedge j) \leq m(i) + m(j)$.

A função-medida em que estamos interessados e que chamaremos agora de medida de conteúdo e denotaremos por 'cont' é definida por

D6-2. $cont(i) =_{DF} m_P(\sim i)$

A partir desta definição, segue-se imediatamente que o valor contido de qualquer E é igual ao valor m_P do correspondente Z.

T6-2. Para cada Z_i, se E_i é $\sim Z_i$, $cont(E_i) = m_P(Z_i)$.

D2 e T11 implicam

T6-3.

a. $cont(i) = 1 - m_P(i)$

b. $m_P(i) = 1 - cont(i)$

c. $cont(\sim i) = m_P(i)$

O seguinte teorema segue de T1 e T3b:

T6-4.

a. $1 \geq cont(i) \geq 0$

b. cont(i) = 0 sse i é L-verdadeira.

c. cont(i) = 1 sse i é L-falsa.

d. $1 > cont\ (i) > 0$ sse i é factual.

e. Se i -L implica j, então $cont(i) \geq cont(j)$.

f. Se i é L-equivalente a j, então $cont(i) = cont(j)$.

g. $cont(i \wedge j) \geq cont(i) \geq cont(i \vee j)$.

h. $cont(i \vee j) = cont(i) + cont(j) - cont(i \wedge j)$.

i. $cont(i \vee j) = cont(i) + cont(j) - 1$ sse i e j são L-exclusivas.

j. $cont(i \wedge j) = cont(i) + cont(j) - cont(i \vee j)$.

k. $cont(i \wedge j) = cont(i) + cont(j)$ sse i e j são L-disjuntas.

l. $cont(\sim i) = 1 - cont(i)$.

m. $cont(i \wedge j) = cont(i) + cont(j)$.

T4*e*, *b* e *c-d* mostram que cont preenche os requisitos de adequação R5-1, R5-2*, e R5-3*, respectivamente.

A condição sob a qual a aditividade é declarada em T4*k* mantida por cont aparece bastante plausível à primeira vista. Se *i* e *j* são L-disjuntas, então os conteúdos de *i* e *j* são exclusivos (T4-2*f*). Nada no que é afirmado por *i* é simultaneamente afirmado por *j*; em outras palavras, não há sentença factual que seja L-implicada tanto por *i* quanto por *j*. No entanto, iremos mais tarde (§ 7) fazer algumas considerações que levantarão algumas dúvidas com respeito a esta condição especial de aditividade.

A medida relativa do conteúdo de *j* em relação a *i* significa o aumento do valor de cont adicionando *j* a *i*. Portanto, em conformidade com o D5-1:

D6-3. $cont(j/i) =_{DF} cont(i \wedge j) - cont(i)$

T6-5.

a. $cont(j/i) = cont(j) - cont(i \vee j)$ (D3, T4j).

b. $\qquad = cont(j)$ sse i e j são L-disjuntas ((a), T4b).

T6-6. $cont(j/i) = cont(i \rightarrow j)$.

Prova: j é L-equivalente a $(i \vee j).(\sim i \vee j)$. Os componentes dessa conjunção são L-disjuntas. Portanto, $cont(j) = cont(i \vee j) + cont(\sim i \vee j)$ (T4k). Portanto, com T5a, $cont(j/i) = cont(\sim i \vee j)$. Mas, $\sim i \vee j$ é L- equivalente a $i \rightarrow j$.

O último teorema é especialmente interessante. Mostra que o conteúdo relativo de *j* em relação a *i* é o mesmo que a medida de conteúdo absoluto da implicação (material) $i \rightarrow j$. Se um receptor "ideal" possui o conhecimento e então adquire o conhecimento j, sua posse de informação só é aumentada na mesma quantidade como se $i \rightarrow j$ fossem adicionados em vez de j. Isso é, de fato, altamente plausível, já que j é uma consequência lógica das sentenças i e $i \rightarrow j$; e um receptor "ideal", por definição, é capaz de tirar tais consequências instantaneamente.

A partir do T6 também vemos que se i L-implica j, $cont(j/i) = 0$. Já sabemos disso visto que vale para todas as nossas *explicata* pela quantidade relativa de informação em virtude de T5-9.

A seguinte inequação, uma consequência imediata de T5a, é de interesse:

T6-7. $cont(j/i) \leq cont(j)$.

Podemos expressar $cont(j/i)$ diretamente em termos de m_P de várias maneiras:

T6-8.

a. $cont(j/i) = m_P(i) - m_P(i \wedge j)$ (D3, T3a).

b. $\quad = m_P(i \wedge \sim j)$ (T6, $i \rightarrow j$ é L-equivalente a $\sim(i \wedge \sim j)$, T3a).

c. $\quad = m_P(i \vee j) - m_P(j)$ (D3, T5a).

Duas sentenças, i e j, que preenchem a condição $m_P(i \wedge j) = m_P(i) \times m_P(j)$, são chamadas indutivamente independentes (ou inicialmente irrelevantes, na terminologia de [Prob.] - p. 356) em relação ao m_P. Nós temos T6-9. Se i e j não têm nenhum predicado primitivo em comum, então

$$m_P(i \vee j) = m_P(i) + m_P(j) - m_P(i) \times m_P(j) \text{ (T1h, D1h)}$$

T6-10.

a. Para qualquer sentença básica B, $m_P(B) = 1/2$.

Prova: $B \vee \sim B$ é L-verdadeira. Portanto, por Tlb, $m_P(B \vee \sim B) = 1$. Daí a afirmação com D1g.

b. Para qualquer conjunção, C_n, de n sentenças básicas com n predicados primitivos distintos, $m_P(C_n) = (1/2)^n$ (D1h, (a)).

c. Se i e i' são isomórficos, então $m_P(i) = m_P(i')$ (D1e).

Agora temos

T6-11. Se i e j não têm nenhum predicado primitivo em comum, então

a. $cont(\sim(i \wedge j)) = cont(\sim i) \times cont(\sim j)$ (Dlh, T3c).

b. $cont(i \vee j) = cont(i) \times cont(j)$.

Prova: $i \vee j$ é L-equivalente $\sim(\sim i. \sim j)$. $\sim i$ e $\sim j$ não têm predicado primitivo em comum desde que i e j também não. Daí a afirmação de (a).

c. $cont(i \wedge j) = cont(i) + cont(j) - cont(i) \times cont(j)$ (T4j, (b)).

Em nosso $\mathcal{L}_3^2$, $cont('Ma \vee Yb') = cont('Ma \vee Ya') = 1/4$ e $cont('Ma \wedge \sim Yb') = 3/4$.

T6-12. Seja D_n uma disjunção de n (≥ 2) sentenças sem predicado primitivo ocorrendo em mais de uma dessas sentenças. Então $cont(D_n)$ = o produto dos valores cont dos n componentes (Tllb).

T6-13.

a. Para qualquer sentença básica B, $cont(B) = 1/2$ (T3a, T10a).

b. Para qualquer disjunção, D_n, de n sentenças básicas com n predicados primitivos distintos, $cont(D_n) = (1/2)^n$ (T3a, T12, (a)).

c. Para qualquer conjunção, C_n, de n sentenças básicas com n predicados primitivos distintos, $cont(C_n) = 1 - (1/2)^n$ (T3a, T10b, (a)).

d. Para qualquer Q-sentença i, $cont(i) = 1 - (1/2)\ ((c)) = 1 - 1/\kappa\ (T2 - lb) = (\kappa - 1)/\kappa$.

$$cont(\text{‘[M}{\sim}\text{Y]a’}) = 3/4 \text{ (Já que } \pi = 2, \kappa = 4 \text{ (T2-1b))}.$$

e. Tendo i a forma $C_1 \vee C_2 \vee ... \vee C_m$ onde cada C é uma conjunção de n sentenças básicas com n predicados primitivos distintos, as mesmas n sentenças atômicas ocorrendo em todas as conjunções. (Nestas circunstâncias, i tem uma forma normal disjuntiva. [Prob.] P. 94 ou qualquer livro didático sobre Lógica Simbólica). Logo

$$cont(i) = 1 - \frac{m}{2^n}.$$

Prova: Quaisquer duas conjunções distintas são L - exclusivas. Portanto, a partir do T4*i*, $cont(i) = cont(C_1) + cont(C_2) + \cdots + cont(C_m) - (m-1)$. Daí a conclusão com (*c*).

$$cont('(\sim Ma \wedge Yb) \vee (Ma \wedge \sim Yb)') = 1 - \frac{3}{2^2} = \frac{1}{4}.$$

Note que esta disjunção é equivalente a 'M*a*VY*b*', isto é, uma disjunção que satisfaz (*b*).

f. Tendo i a forma $\mathrm{D}_1 \wedge \mathrm{D}_2 \wedge \ldots \wedge \mathrm{D}_m$ onde cada D é uma disjunção de n sentenças básicas com n predicados primitivos distintos, as mesmas n sentenças atômicas ocorrendo em todas as disjunções. (Nestas circunstâncias, i tem forma normal conjuntiva. [Prob.] P. 95.)

Então

$$cont(i) = \frac{m}{2^n}$$

Prova: Quaisquer duas disjunções distintas são L-disjuntas. Portanto, a partir de T4k, $cont(i) = cont(\mathrm{D}_1) + cont(\mathrm{D}_2) + \cdots + cont(\mathrm{D}_m)$ Daí a afirmação com (b).

$$\mathrm{cont}('(\mathrm{Ma} \vee \mathrm{Yb}) \wedge (\sim\mathrm{Ma} \vee \mathrm{Yb}) \wedge (\mathrm{Ma} \vee \sim\mathrm{Yb})') = \frac{3}{2^{\mathrm{n}}} = \frac{3}{4}$$

Observe que essa conjunção é L-equivalente a 'Ma. Yb ', isto é, uma conjunção

satisfazendo (c).

T6-14. Se i e i 'são isomórficas e j e j' são isomórficas com base na mesma permutação das constantes individuais, então $cont(j'/i') = cont(j/i)$ (T8b, T10a).

T6-15.

a. Para quaisquer duas sentenças básicas, B_i e B_j, com diferentes predicados primitivos, $cont(B_j/B_i) = 1/4\ (T13c, T10a) = 1/2\ cont(B_i)$ (T13a).

$$\mathrm{cont}(\text{‘Ya’}/\text{‘Ma’}) = 1/4.$$

b. Seja $B_1, B_2, \ldots, B_n$ sentenças básicas com n predicados primitivos distintos. Seja C_m a conjunção do primeiro m deles. Então, para todo m $(m = 2, \ldots, n - l)$,

$$\mathrm{cont}(B_{m+1}/C_m) = \frac{1}{2^{m+1}}.$$

Prova: $C_m \wedge B_{m+1} = C_{m+1}$, consequentemente

$$\mathrm{cont}\left(\frac{B_{m+1}}{C_m}\right) = cont(C_{m+1}) - cont(C_m) = 1 - \frac{1}{2^{m+1}} - \left(1 - \frac{1}{2^m}\right)(T13c) = \frac{1}{2^{m+1}}$$

T6-16. Seja i e j sentenças moleculares sem predicado primitivo em comum. Então $cont(j/i) = cont(j) - cont\ (i) \times cont(j)\ (Tllc) = cont(j) \times (1 - cont(i)) = cont(j) \times cont(\sim i)(T41) = cont(j) \times m_P(i)\ (T3c)$.

§ 7. O segundo explicatum: Medida de informação (inf)

Um dos teoremas derivados na seção anterior afirma que se i e j são sentenças básicas com diferentes predicados primitivos, então $cont(j/i) = 1/2cont(i)$. Já que frases básicas com diferentes predicados primitivos são indutivamente independentes, este resultado faz com que o cont pareça inadequado como um *explicatum* para in. Acontece que nenhum *explicatum* preenchendo todos os nossos requisitos intuitivos para uma função de quantidade de informação é possível, indicando certa inconsistência entre estes requisitos. Cont preenche um conjunto parcial desses requisitos, e um diferente, embora sobreposto, conjunto parcial é preenchido por outra função, chamada medida de informação, denotada por 'inf', e definida como

$$\inf(i) = Log\frac{1}{1 - cont(i)}.$$

É mostrado que

$$\inf(h, e) = Log\frac{1}{c_P(h, e)}$$

onde $c_P(h, e)$ é o grau de confirmação da hipótese h sobre a evidência e, definido como

$$\frac{m_P(e, h)}{m_P(e)}.$$

O último teorema da seção anterior (T6-15) pode não aparecer inteiramente plausível. De acordo com este teorema, se um receptor "ideal" sem conhecimento prévio recebe uma sequência de n sentenças básicas com n diferentes predicados primitivos, a quantidade de informação que ele recebe da primeira frase é 1/2, da segunda apenas 1/4, a partir do terceiro 1/8, de cada metade, tanto quanto do precedente. E este será o caso, apesar do fato de que essas sentenças básicas são independentes umas das outras não só dedutivamente, mas também indutivamente. Tem-se a sensação de que sob tais condições a quantidade de informação contida em cada sentença não deva depender de ser precedida por outra de seu tipo.

Uma inconsistência em nossas intuições, a qual já sugerimos acima (§6), torna-se agora ainda mais proeminente. O sentimento a que nos referimos no parágrafo anterior pode ser expresso também como um requisito que a aditividade deve manter para a quantidade de informação transportada pela conjunção de duas sentenças, se essas sentenças forem indutivamente independentes. Vimos, no entanto, que a aditividade vale apenas para cont se essas sentenças forem L disjuntas e não tenham conteúdo em comum. Agora, está claro que duas sentenças, B_1 e B_2, com diferentes predicados primitivos, têm conteúdo em comum: A sentença factual $B_1 \vee B_2$, por exemplo, é L-implicada por cada uma. No entanto, esta condição de aditividade pareceu plausível em seu contexto.

Parece melhor resolver este conflito de intuições assumindo que não há um *explicandum* "quantidade de informação semântica", mas pelo menos dois, para um dos quais cont é de fato, um *explicatum* adequado, enquanto o *explicatum* para o outro ainda tem que ser encontrado.

Vamos agora declarar o requisito adicional de maneira formal:

R7-1 Se i e j são indutivamente independentes, então $in(i \wedge j) = in(i) + in(j)$.

De Ri e D5-1 segue imediatamente:

T7-1 Se B_i e B_j são duas sentenças básicas com predicados primitivos distintos, então

$$in(B_j/B_i) = in(B_j)$$

Vamos também decidir, em nome da normalização, atribuir a cada sentença básica um valor de 1.

R7-2. Para qualquer sentença básica B, in(B) = 1.

Temos agora

T7-2. Para uma conjunção de n sentenças básicas, C_n, com n predicados primitivos distintos, em $in(C_n) = n$ (R1, R2).

T6-13c afirmou que $cont(C_n) = 1 - (1/2)^n$, portanto

$$2^n = \frac{1}{1 - cont(C_n)}$$

consequentemente

$$n = Log\frac{1}{1 - cont(C_n)}$$

(onde 'Log' é a abreviação de 'logaritmo na base 2'). Isso, combinado com os rendimentos T2.

T7-3. Para uma conjunção de n sentenças básicas, C_n, com n predicados primitivos distintos,

$$in(C_n) = Log\frac{1}{1 - cont(C_n)}$$

T3 nos dá a liderança para definir o segundo *explicatum* para "quantidade de informação". Essa nova função será chamada de medida de informação e denotada por 'inf'. Estendendo a relação declarada em T3 para manter por todas as sentenças, definimos

D7-1 Para qualquer sentença *i*,

$$\inf(i) = Log\frac{1}{1 - cont(i)}$$

D1 pode ser utilmente transformado em

T7-4.

a. $inf(i) = -Log(1 - cont(i)) = -Log\ cont(\sim i) = Log\frac{1}{cont(\sim i)}$

b. $inf(i) = -Log\ cont(i)$.

c. $cont(\sim i) = 2^{-\inf(i)}$

d. $cont(i) = 1 - 2^{-\inf(i)}$

T7-5.

a. $= Log\frac{1}{m_P(i)}$ (D1, T6-3)

b. $= -Log\ m_P(i)$

A forma de T5*a* é análoga à definição habitual de quantidade de informação na teoria da comunicação. No lugar do conceito de probabilidade no sentido estatístico (frequência relativa) usado naquela definição, temos aqui a probabilidade lógica (indutivo). Para uma discussão detalhada da relação entre esses dois conceitos, veja [Prob.] §§3, 10.

T7-6. $m_P(i) = 2^{-\inf(i)}$

Uma série de outros teoremas para inf pode ser facilmente derivada. Vamos mencionar apenas alguns deles.

T7-7. $\inf(\sim i) = Log\frac{1}{1-m_P(i)}$ (D1, T6-3) $= -Log\ (1 - m_P(i))$.

T7-8.

a. $0 \leq \inf(i) \leq \infty$(T6-4a)

b. $inf(i) = 0$ sse i é L-verdadeira (T6-4*b*).

c. $inf(i) = \infty$ sse i é L-falsa (T6-4*c*).

d. $inf(i)$ é positivo finito sse i é factual (T6-4*d*).

e. Se i L-implica j, então $inf(i) \geq inf(j)$ (T6-4*e*).

f. Se i é L-equivalente a j, então $inf(i) = inf(j)$ (T6-4*f*).

g. $inf(i \wedge j) \geq inf(i) \geq inf(i \vee j)$ (T6-4*g*).

h. $\inf(i \wedge j) = -Log\ cont(\sim i \vee \sim j)(T4a)$

$$= -Log\ (cont(\sim i) + cont(\sim j) - cont(\sim i \wedge \sim j)\ \text{(T6-4h)}$$

$$= -Log\ 2^{-\inf(i)} + 2^{-\inf(j)} - 2^{-\inf(i \vee j)}\ \text{(T4c).}$$

i. Se i e j são L-disjuntas, então $inf(i \wedge\ j) = -Log\left(2^{-\inf(i)} + 2^{-\inf(j)} - 1\right)$ ((*h*), (*b*)).

j. $inf(i \vee j)\ =\ -Log\ cont\ (\sim i \wedge\ \sim j)$ (T4*a*) $= -Log\left(1 - 2^{-\inf(\sim i \wedge \sim j)}\right)$ (T4*d*).

k. Se i e j são L - exclusivas (daí ~i e ~j L-disjuntas), então

$$\inf(i \vee j) = -Log\big(cont(\sim i) + cont(\sim j)\big)\big((j), T4k\big) = -Log\left(2^{-\inf(i)} + 2^{-\inf(j)}\right)(T4d)$$

1. inf (~ i) = inf (i) = $-Log\left(2^{\inf(i)} - 1\right)$ (T4b, d).

Considerando que a correspondência entre T8*a* a *g* e T6-4*a* até *g* é direta, T8*h* através de *l* é muito mais complicada e muito menos conveniente para cálculo do que seus teoremas correspondentes T6-4*h* até *l*.

Em contraste com a complicada fórmula T8*i*, temos, no entanto,

T7-9 Aditividade. $inf(i \wedge j) = inf(i) + inf(j)$ sse i e j sejam indutivamente independentes.

Prova: $inf(i \wedge j) = -Log\ m_P(i \wedge j)(T5b)$

$$= \ -Log(m_P(i) \times m_P(j))\ \text{(por hipótese)}$$

$$= \ -Log\left(2^{\inf(i)} \times 2^{-\inf(j)}\right)(T6)$$

$$= Log\ 2^{\inf(i)} + 2^{\inf(j)} = \inf(i) + \inf(j)$$

Para T6-13 corresponde

T7-10

a. Para qualquer sentença básica B, $inf(B) = 1$.

b. Para qualquer disjunção, D_n, de n sentenças básicas com n predicados primitivos distintos,

$$inf(D_n) = Log\frac{1}{1-\left(\frac{1}{2}\right)^n} = Log\frac{2^n}{2^n-1} = n - Log(2^n - 1).$$

c. Para qualquer conjunção, $\mathrm{C_n}$, de n sentenças básicas com n predicados primitivos distintos, $\mathrm{inf}(\mathrm{C_n}) = n$.

d. Para qualquer Q- sentença i, $inf(i) = \pi$

e. Tendo i forma normal disjuntiva: $\mathrm{C_1} \vee \mathrm{C_2} \vee \ldots \vee \mathrm{C_m}$. Seja todo C uma conjunção de n sentenças básicas com n predicados primitivos distintos, as mesmas n sentenças atômicas que ocorrem em todas as conjunções. Então $inf(i) = n - Log\ m$.

$$inf\ ('(Ma \wedge Yb) \vee (\sim Ma \wedge Yb) \vee (Ma \wedge \sim Yb)') = 2 - Log\ 3 (= 0{,}412).$$

f. Seja i tendo a forma normal conjuntiva: $\mathrm{D_1} \wedge \mathrm{D_2} \wedge \ldots \wedge \mathrm{D_m}$, sendo todo D uma disjunção de n sentenças básicas com n predicados primitivos distintos, as mesmas n sentenças atômicas ocorrendo em todas as disjunções. Então $inf(i) = n - Log\ (2^n - m)$.

$$inf\ ('(Ma \vee Yb) \wedge (\sim Ma \vee Yb) \wedge (Ma \vee \sim Yb)') = 2 - Log\ (2^2 - 3) = 2$$

T8*e*, *b* e *d* mostram que inf preenche R5-l a R5-3. T9 corresponde a R1 e Tl0*a* a R2. Assim, é mostrado que inf preenche todos os nossos requisitos para o segundo *explicatum* para quantidade de informação.

A tabela a seguir fornece inf-valores aproximados para $\mathrm{D_2}$ (Tl0*b*) até $\mathrm{D_{10}}$:

	Tabela II	
n		$Inf(D_n)$
2		0.412
3		0.192
4		0.093
5		0.046
6		0.023
7		0.0113
8		0.0056
9		0.0028
10		0.0014

Definimos agora a medida relativa de informação da maneira familiar:

D7-2. $\mathrm{inf}(j/i) =_{DF} \mathrm{inf}(i \wedge j) - \mathrm{inf}\ (i)$

T7-11

a. Para quaisquer duas sentenças básicas, B_i e B_j, com predicados primitivos distintos, inf$(B_j/B_i) = 1$ (D2, T10c, a) =inf(B_i) (T10a)

$inf('Ma'/'Yb') = inf('Ma'/'Ya') = 1.$

b. Sejam $B_1, B_2, \dots, B_n$ sentenças básicas com n predicados primitivos distintos.

Seja C_m a conjunção dos primeiros m deles. Então, para todo m ($m = 2, .., n - l$),

$$\text{inf}\ (B_{m+1}/C_m) = 1 \text{ (D2, T10c, (a))}$$

T7-12.

a. $inf(j/i) = inf(j)$ sse i e j são indutivamente independentes (D2, T9).

b. Se i e j não possuem predicados primitivos em comum, $inf(j/i) = inf(j)$ (T6-9*a*, (*a*)).

Em [Prob.] § 55, o conceito de grau de confirmação de uma hipótese h sobre a evidência e, com base em uma dada faixa de medida m, é definido como segue:

$$c(h, e) = \frac{m(e \wedge h)}{m(e)}$$

e L-implica h se, e somente se, o alcance de e está totalmente contido no alcance de h. E se, no entanto, apenas uma parte de R (e) está contida em R (h), então nenhuma das relações costumeiras da lógica dedutiva se mantém entre e e h. Se, digamos, aquela parte de R (e) que está contida em R (h) é três quartos de R (e), medido por m, se, em outras palavras,

$$\frac{m(e \wedge h)}{m(e)} = \frac{3}{4}$$

então diremos que a hipótese h é confirmada pela evidência e até o grau $3/4$ e escrevemos essa relação, que é fundamental para a lógica indutiva, como '$c(h, e) = 3/4$'. c significa um *explicatum* para probabilidade indutiva (relativa).

A figura 1 pode ser de alguma ajuda para uma visualização da diferença entre L - implicação e grau de confirmação como dependente das relações entre os alcances da hipótese e da evidência.

Para uma função m_P, temos mais especificamente

D7-3. $c_P(h, e) = \frac{m(e \wedge h)}{m(e)}$

T7-13. Se inf e c_P são baseados no mesmo m_P, então

$$Inf(h/e) = Log\ \frac{1}{c_P(h,e)} = -Log\ c_P(h,e)$$

Prova: $\inf(h/e) = \inf(e \wedge h) - \inf(e) = Log\ m_P(e) - Log\ m_P(e \wedge h) = Log\ \frac{m(e)}{m(e \wedge h)} = Log\ \frac{1}{c_P(h,e)}$.

Este teorema mostra a forte conexão que existe entre a medida relativa de informação de uma nova mensagem *h* com relação ao conhecimento *e* e o grau de confirmação de uma hipótese *h* sobre a evidência *e*, em outras palavras, a probabilidade relativa indutiva de uma hipótese *h* sobre a evidência *e*. A caracterização de *h* como mensagem e como conhecimento, por um lado, ou como hipótese e evidência, por outro, tem valor didático apenas; *h* e *e*, estritamente falando, são simplesmente quaisquer sentenças do dado sistema. T13 mostra que $inf(h/e)$ é maior quanto mais improvável estiver na evidência *e*. Que a quantidade relativa de informação transportada por uma sentença deve aumentar

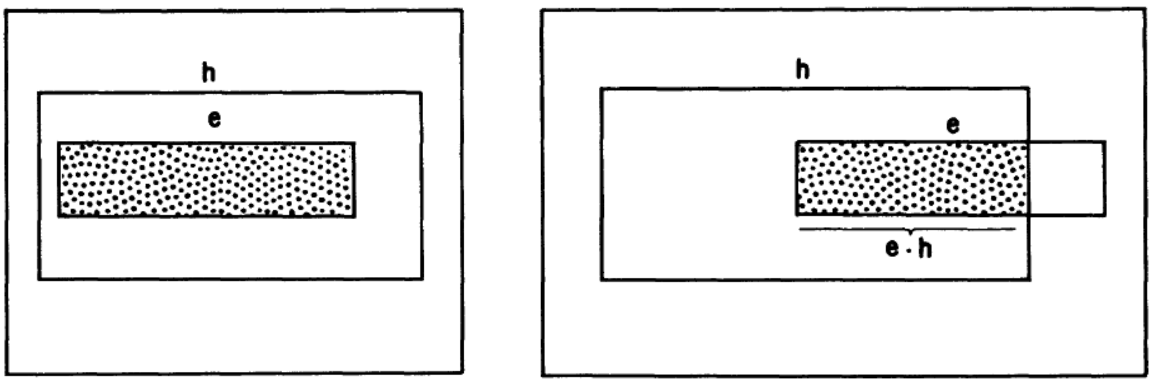

Fig. 1

Lógica Dedutiva
'e L-implica h' significa que o alcance de e está inteiramente contido no de h.

Lógica Indutiva
'$c(h,e) = 3/4$' significa que três quartos do alcance de e está contido no de h

com seu grau de improbabilidade parece plausível. Isso vale também para cont, se *e* permanece fixo, como mostrado pelo seguinte teorema.

T7-14. Se cont e C_P são baseados no mesmo m_P, então

$$cont(h/e) = m_P(e) \times \left(1 - C_P(h,e)\right) = m_P(e) \times c_P(\sim h, e)$$

Prova: $cont(h/e) = m_P(e) - m_P(e \wedge h)$ (T6-8a)

$$= m_P(e) \times \frac{m_P(e) - m_P(e \wedge h)}{m_P(e)}$$

$$= m_P(e) \times \left(1 - c_P(h,e)\right) \text{ (D3).}$$

Observe, no entanto, que para variável e não precisa ser o caso de que o menor $c_P(h,e)$ é, o maior $cont(h/e)$ será por causa do fator $m_P(e)$. (Veja o final do §10, abaixo).

§8. Comparação entre cont e inf

As medidas cont e inf são comparadas em maior detalhe. Ambas exibem propriedades que parecem intuitivamente plausíveis e outras que parecem intuitivamente implausíveis. A comparação formalmente mais impressionante é dada pelo seguinte par de teoremas:

$$cont(h/e) = m_P(e) - m_P(e \wedge h),$$
$$inf(h/e) = Log\, m_P(e) - Log\, m_P(e \wedge h).$$

Agora estamos prontos para uma comparação entre as duas *explicata*s para quantidade de informação. Vamos começar com a afirmação de alguns teoremas correspondentes um ao lado do outro para um melhor confronto.

T6-4k. $cont(i.j) = cont(i) + cont(j)$ sse i e j são L-disjuntas.	T7-9. $inf(i.j) = inf(i) + inf(j)$ sse i e j são indutivamente independentes.
T6-4m. $cont(i.j) \leqslant cont(i) + cont(j)$.	
T6-13a. Para qualquer sentença básica B, $cont(B) = 1/2$.	T7-10a. Para qualquer sentença básica B, $inf(B) = 1$.
T6-13c. Para qualquer conjunção, C_n, de n sentenças básicas com n predicados primitivos distintos, $cont(C_n) = 1 - (1/2)^n$.	T7-10c. Para qualquer conjunção, C_n, de n sentenças básicas com n predicados primitivos distintos, $inf(C_n) = n$.
T6-15a. Para quaisquer duas sentenças básicas, B_i e B_j, com predicados primitivos distintos, $cont(B_j/B_i) = 1/4 = 1/2\ cont(B_i)$.	T7-11a. Para quaisquer duas sentenças básicas, B_i e B_j, com predicados primitivos distintos, $inf(B_j/B_i) = 1 = inf(B_i)$.
T6-15b. Sejam $B_1, B_2, \ldots, B_n$ sentenças básicas com n predicados primitivos distintos. Seja C_m a conjunção do primeiro m deles. Então, para cada m $(m = 2, \ldots, n-1)$, $cont(B_{m+1}/C_m) = \frac{1}{2^{m+1}}$.	T7-11b. Sejam $B_1, B_2, \ldots, B_n$ sentenças básicas com n predicados primitivos distintos. Seja C_m a conjunção do primeiro m deles. Então, para cada m $(m = 2, \ldots, n-1)$, $inf(B_{m+1}/C_m) = 1$.
T6-6h. $cont(j/i) = cont(j)$ sse i e j são L - disjuntas.	T7-12a. $inf(j/i) = inf(j)$ sse i e j são indutivamente independentes.
T6-7. $cont(j/i) \leqslant cont(j)$.	
T6-4ℓ. $cont(\sim i) = 1 - cont(i)$.	T7-8ℓ. $inf(\sim i) = inf(i) - Log(2^{inf(i)} - 1)$.

Vemos que as condições de aditividade para cont e inf são totalmente diferentes. Esta divergência não é de todo surpreendente. Pelo contrário, insatisfação com a condição de aditividade estabelecida para cont em T6-4k foi uma das razões para a nossa busca por outro *explicatum* da quantidade de informação. É de algum interesse psicológico notar que o senso comum provavelmente preferiria T7-9 a T6-4k, enquanto inf não tem propriedade comparável àquela exibida por cont em T6-4m, um teorema que parece altamente intuitivo.

A contra-intuição da falta de uma contrapartida para T6-4m pode ser reduzida pelo exemplo a seguir. Considere um sistema com 6 predicados primitivos, a , portanto, com = 64 Q-propriedades. Todas as funções m próprias têm valores iguais para o 64 Q-sentenças com a mesma constante individual, daí o valor 1/64 para cada uma. Seja i ' a'. é a disjunção dos primeiros 32 Q's. Por isso, (i) = 1/2. Assim sendo $inf(i) = -Log(1/2) = 1$.

Seja M uma disjunção de 32 Q's, isto é, de e os últimos 31 Q's Seja j 'Ma'. Então $m(j) = 1/2$ e $inf(j) = 1$. $i \wedge j$ é equivalente a L para; consequentemente, é uma sentença muito forte. $m(i \wedge j) = 1/64. inf(i \wedge j) = -Log(1/64) = 6$. Isto é três vezes tanto quanto a soma dos valores de inf dos dois componentes. Este resultado torna-se plausível se percebermos que i diz apenas que tem um de certos 32 Q's, mas que pela adição de j, que por si só também não diz mais do que a que tem um de 32 Q's, nossa informação sobre a situação é imediatamente específica; isto é, é especificada como dizendo que a tem um Q particular.

Continuando a comparação, podemos descartar a diferença entre T6-13a e T7-10a como não essencial, o número 1 em T7-10a é apenas uma questão de normalização.

No entanto, as diferenças entre T6-13c e T7-10c, T6-15a e T7-11a, e T6-15b e T7-1lb são decisivas. Considerando que o valor contido de uma sentença básica para uma conjunção de sentenças básicas com diferentes predicados primitivos é sempre menos que o seu cont-valor absoluto e diminui, além disso, com o número de conteúdos na conjunção, o inf-valor de uma sentença básica relativa a tal conjunção é igual ao seu cont- valor absoluto e, portanto, também é independente do número de componentes nesta conjunção.

A relação entre cont e inf é trabalhada no talvez mais simples e mais marcante modo pelo seguinte par de fórmulas que aparecem nas provas de T7-13 e T7-14:

$$\text{inf}(h/e) = Log\ m_P(e) - Log\ m_P(e \wedge h). \qquad (1)$$

$$cont(h/e) = m_P(e) - m_P(e \wedge h). \qquad (2)$$

Para a evidência tautológica, obtemos

$$inf(h/t) = inf(h) = -Log\ m_P(h) \qquad (3)$$

e

$$cont(h/t) = cont(h) = 1 - m_P(h) \qquad (4)$$

fórmulas que são nada mais do que variantes de T7-5*b* e T6-3*a*, mas veja agora muito mais afim, especialmente se escrevermos (3) como

$$inf(h/t) = inf(h) = Log\ \ 1 - Log\ m_P(h) \qquad (3')$$

Vamos ilustrar a relação entre cont e inf também no seguinte numérico exemplo.

Sejam B_1, B_2, ... sentenças básicas com predicados primitivos distintos.

Seja C_1, B_1, C_2 seja $B_1 \wedge B_2$, ..., C_n seja B_1, B_2, ... B_n, ... Então cont e inf têm os seguintes valores para estes C, de acordo com T6-13c e T7-10c:

	Tabela III	
$\boldsymbol{C_i}$	cont (C_i)	inf (C_i)
$\boldsymbol{C_1}$	1/2	1
$\boldsymbol{C_1}$	3/4	2
$\boldsymbol{C_1}$	7/8	3
$\boldsymbol{C_1}$	15/16	4
.	.	.
.	.	.
.	.	.
$\boldsymbol{C_1}$	(n-1)/n	n
.	.	.
.	.	.
.	.	.

§9. Funções-D e funções-I

> Nem todas ases funções-m_P podem ser consideradas *explicatas* igualmente adequadas de probabilidade indutiva. Parece que somente aquelas que preenchem o requisito adicional de relevância instancial – funções-m_I - são adequadas para propósitos científicos comuns, enquanto que a função-m_P que exibe irrelevância instancial - m_D - possui propriedades que a tornam adequada para situações em que o raciocínio indutivo é de menor importância. Computações com m_D e in-funções baseadas nelas são relativamente fáceis devido ao fato de que m_D atribui valores iguais a todas as descrições-de-estado. Uma consequência é, por exemplo, que uma condição suficiente para $m_D(i \wedge j) = m_D(i) \times m_D(j)$ é que i e j não devam ter sentenças atômicas em comum, enquanto apenas a condição muito mais forte que i e j não devam ter predicados primitivos em comum seja suficiente para o teorema correspondente a m_I.

Nem todo m_P pode servir de base para um método indutivo que esteja de acordo com os procedimentos científicos habituais (*cf*. [Cont.] §2). Há pelo menos um requisito adicional para a adequação de uma função-m para servir como um *explicatum* para (absoluto, inicial) probabilidade indutiva. Isto é

R9-1 Requisito de relevância instantânea. Seja 'M' um predicado molecular e factual. Seja e qualquer sentença molecular não-L-falsa. Sendo i e h sentenças completas de 'M' comduas constantes individuais distintas que não ocorrem em e. Então

$$\frac{m(e \wedge i \wedge h)}{m(e \wedge i)} > \frac{m(e \wedge h)}{m(e)}$$

(Isto pode ser formulado mais simplesmente em termos de 'c' como

$$c(h, e \wedge i) > c(h, e).)$$

O requisito diz, com efeito, que uma instância de uma propriedade é positivamente relevante para (a previsão de) outra instância da mesma propriedade. Isso parece característica básica de todo raciocínio indutivo relativo à previsão de um evento futuro.

Nós, portanto, definimos a função m indutiva (no sentido mais restrito), a ser denotada por ' m_I' como

D9-1 m é uma m-função indutiva $=_{DF}$ m é um m_P e cumpre R1.

Entre as funções m próprias que não cumprem R1, há uma que cumpre, por assim dizer, uma exigência de irrelevância instancial. Para esta função m, ser denotado por 'm_D' ('D' para 'dedutivo', uma vez que esta função desempenha um papel especial na lógica dedutiva), exemplos observados de uma propriedade molecular não influenciam a previsão de futuras instâncias dessa propriedade. A experiência não pode nos ensinar nada sobre o futuro se essa função é aplicada. Tem, no entanto, grande importância: sua definição é de extrema simplicidade, os cálculos em sua base são relativamente fáceis, e os resultados obtidos pelo seu uso podem ter pelo menos valor aproximado nos casos em que a experiência é estimada como sendo de pouca ou nenhuma influência.

A definição de m_D incorpora um princípio que parece muito plausível o senso comum não treinado, ou seja, o princípio de atribuição de m-valores iguais a todas as descrições-de-estado. É de algum interesse psicológico que esse procedimento bastante óbvio leve a um método indutivo que é inaceitável como método final. (A função designada aqui por ' m_D' foi denotada por '$m\dagger$' em [Prob.] §100A e por 'm_∞' em [Cont.] §13.)

Definimos

D9-2.

a. Para cada $Z_i, m_D(Z_i) =_{DF} 1/z$

b. Para cada sentença L-falsa j, $m_D(j) =_{DF} 0$

c. Para cada sentença não-L-falsa j, $m_D(j) =_{DF}$ a soma dos m_D-valores para o Z em R(j); isto é $r(j)/z$, onde r (j) é o número das descrições-de-estado em R(j).

Pode ser facilmente verificado que m_D preenche as condições D6-l*a* a D6-l*g* e D6-li. Que m_D também cumpre D6-l*h* e é, portanto, uma função m_D, segue a partir do teorema muito mais forte T9-1 abaixo.

T9-1 Se i e j não têm sentenças atômicas em comum, então

$$m_D(i \wedge j) = m_D(i) \times m_D(j).$$

Prova: Seja K_1 a classe das sentenças atômicas que ocorrem em i, K_2 a classe dessas sentenças atômicas que ocorrem em j, K_3 a classe de todos as outras sentenças atômicas. Seja C_1 a classe daquelas conjunções que contêm, para cada sentença atômica em K_1, seja ela ou a sua negação, mas não ambas, nem qualquer outro componente. Sejam C_2 e C_3 determinados analogamente em relação a K_2 e K_3. Seja c_1 o número das conjunções em C_1. Seja c_2 e c_3 determinados analogamente com relação a C_2 e C_3. (Entretanto, se C_3 for vazio, seja c_3= 1.)

Cada Z é uma conjunção de três conjunções (desconsiderando a ordem) pertencentes respectivamente a C_1, C_2, e C_3. Assim sendo

$$z = c_1 \times c_2 \times c_3. \qquad (1)$$

Seja $c_1(i)$ o número dessas conjunções em C_1 as quais L-implica i, e seja c_2(j) o número dessas conjunções em C_2 as quais L-implica j. (Observe que i não pode ser L-implicada por qualquer conjunção de C_2 ou C_3, nem j pode ser L-implicada por qualquer conjunção de C_1 ou C_3.) Portanto

$$r(j) = c_1(j) \times c_2 \times c_3. \qquad (2)$$

e

$$r(j) = c_2(j) \times c_1 \times c_3. \qquad (3)$$

Mas, pela mesma razão, também temos

$$r(i \wedge j) = c_1(i) \times c_2(j) \times c_3. \qquad (4)$$

De (2) e (3) obtemos

$$\begin{aligned} r(i) \times r(j) &= c_1(i) \times c_2 \times c_3 \times c_2(j) \times c_1 \times c_3 \\ &= r(i \wedge j) \times c_1 \times c_2 \times c_3 \text{ (de (4))} \\ &= r(i \wedge j) \times z \text{ (de (1))} \end{aligned} \qquad (5)$$

Dividindo por z^2 finalmente conseguimos

$$\frac{r(i)}{z} \times \frac{r(j)}{z} = \frac{r(i \wedge j)}{z} \qquad (6)$$

do qual a afirmação se segue, por D2*c*.

Como m_D é uma função m_P, todos os teoremas declarados em §6 para m_P sustentam também para m_D. Mas alguns deles, tendo forma condicional, podem ser fortalecidos pelo enfraquecimento do antecedente. Nós temos, por exemplo, em analogia ao T6-9,

T9-2. Se i e j não têm sentença atômica em comum, então

$$m_D(i \vee j) = m_D(i) + m_D(j) - m_D(i) \times m_D(j)$$

e em analogia a T6-10b,

T9-3 Para qualquer conjunção, C_n, de n sentenças básicas com n sentenças atômicas distintas, $m_D(C_n) = (1/2)^n$.

Para essa função cont com base em m_D de acordo com D6-2, $cont_D$, temos

T9-4

a. Para cada E_i, $cont_D(E_i) = 1/z$

b. Para cada sentença j, $cont_D(j) = n/z$, onde n é o número de E que pertence a Cont(j).

$cont_D$ tem vantagens e desvantagens semelhantes às de m_D. Tlb aponta para a extrema simplicidade, pelo menos em princípio, de sua computação.

Todos os teoremas sobre cont declarados no §6 também sustentam, é claro, para $cont_D$ e todo $cont_I$, as cont-funções definidas com base no m_I, analogamente a D3. Com relação ao $cont_I$, nenhum teorema adicional na forma de igualdade pode ser derivado de Ri. Não nos preocuparemos em derivar algumas desigualdades de Ri e dos teoremas anteriores, especialmente porque trataremos mais adiante (§10), com alguma extensão, um exemplo numérico baseado em uma função m_I específica.

No que diz respeito a $cont_D$, no entanto, vários teoremas aplicados a $cont_P$ podem ser reforçados enfraquecendo a condição no antecedente, em completa analogia a relação entre m_D e m_P. T6-9, T6-10b, T6-11, T6-12, T6-13b, c, e, f, T6-15 e T6-16 aplica a $cont_D$, mesmo que a expressão 'predicado (s) primitivo (s)' em seus antecedentes seja substituída por "sentença(s) atômica(s)". Que isso seja assim é plausível em vista do T1. Mas também é fácil verificar a verdade de nossa afirmação geral inspecionando as provas desses teoremas.

Vamos afirmar, no entanto, também um teorema que não é uma contrapartida de um teorema anterior:

T9-5

a. Para cada conjunção, C_n, de n E distinto, $cont_D(C_n) = n/z$

b. Para cada conjunção, C_n, de n E distinto, diferente de E_i,

$$cont_D(E_i/C_n) = 1/z \text{ (D6-2, } (a)\text{)}.$$

A relação entre inf_D, definida com base em $cont_D$ seguindo D7-1, e inf é o mesmo que entre $cont_D$ e cont. Portanto, devemos declarar apenas aqueles teoremas que são baseados em T4 e T5.

T9-6.

a. Para cada E, $inf_D(\mathrm{E}) = \beta - \mathrm{Log}\,(z-1)$

Prova: $inf_D(E) = Log \frac{1}{1-\frac{1}{z}} (T4a) = Log \frac{z}{z-1} = \beta - Log\,(z-1)\ (TZ\text{-}lc)$.

b. Para cada conjunção, C_n , de n E distinto, $\mathrm{inf}_D(C_n) = \beta - \mathrm{Log}\,(z-n)$

Prova: $\mathrm{inf}_D(C_n) = \mathrm{Log\ n} \frac{1}{1-\frac{n}{z}} (T5a) = \mathrm{Log} \frac{z}{z-n} = \beta - \mathrm{Log}\,(z-n)$ (T2-lc).

c. Para cada conjunção, C_n, de n E distinto, diferente de E_i,

$\mathrm{inf}\,(E_i/C_n) = Log(z-n) - Log(z-n-1)\ (D7\text{-}2, (b))$

De acordo com o correlato de T7-10*e*, $inf_D(i) = n - Log\ m$, onde i tenha forma normal disjuntiva: $C_1 \vee C_2 \vee \ldots \vee C_m$ sendo cada C uma conjunção de sentenças básicas com *n* sentenças atômicas distintas, as mesmas *n* sentenças atômicas ocorrendo em todas as conjunções. De acordo com um teorema conhecido no cálculo sentencial, existe para cada sentença molecular uma sentença em forma normal disjuntiva L- equivalente a ela (ver, por exemplo, [Prob.] D21-2). Segue-se que para cada sentença molecular *i*, $inf_D(i)$ tem a forma $n - Log\ m$, onde tanto *n* como *m* são inteiros. Por isso, é fácil calcular o valor inf_D de qualquer sentença molecular. Tal sentença tem que ser transformada em uma de suas formas normais disjuntivas, de acordo com algum procedimento padrão disponível para este fim. Então o número de seus componentes deve ser contado, assim como o número de sentenças atômicas em um desses componentes. Finalmente, uma tabela para Log *m*, para o inteiro *m*, terá que ser consultada e uma simples subtração executada. Para fins de referência, essa tabela é fornecida aqui para alguns valores integrais selecionados de *m*.

Tabela IV					
m	Log m	m	Log m	m	Log m
1	0.0000	39	5.2853	57	5.8328
2	1.0000	40	5.3219	58	5.8579
3	1.5849	41	5.3575	59	5.8826
4	2.0000	42	5.3923	60	5.9068
5	2.3219	43	5.4262	61	5.9307
6	2.5849	44	5.4594	62	5.9541
7	2.8073	45	5.4918	63	5.9772
8	3.0000	46	5.5235	64	6.0000
9	3.1699	47	5.5545	100	6.6438
10	3.3219	48	5.5849	128	7.0000
16	4.0000	49	5.6147	250	7.9657
32	5.0000	50	5.6438	251	7.9715
33	5.0443	51	5.6724	252	7.9772
34	5.0874	52	5.7004	253	7.9829
35	5.1292	53	5.7279	254	7.9886
36	5.1699	54	5.7548	255	7.9943
37	5.2094	55	5.7813	256	8.0000
38	5.2479	56	5.8073	1000	9.9657

Seja o E em nossos $\mathcal{L}_3^2$, $E_1, E_2, \ldots, E_{64}$. Seja C_m a conjunção do primeiro m E. Então, $cont_D(C_m) = m/64$ (T2a) e $\inf_D(C_m) = 6 - Log(64 - m)$ (T3b). Tabela V dá os valores relativos e absolutos de cont D para os primeiros seis valores de m e para os últimos seis valores de m.

Vemos a partir disso que, se uma série de mensagens é recebida, cada uma sendo E, então $cont_D$ cresce por cada uma dessas mensagens pelo mesmo valor, ou seja, por $1/64$, de 0 a 1. $\inf_D$, no entanto, se comporta de maneira diferente. Ela cresce de 0 a ∞ em quantidades desiguais. A primeira mensagem contribui apenas com uma pequena fração. Cada nova mensagem contribui um pouco mais do que a anterior. A anterior a antepenúltima contribui com menos de $1/2$. A antepenúltima contribui com mais de $1/2$. A penúltima contribui com 1. E a última mensagem contribui com ∞. Este comportamento do $\inf_D$ torna-se plausível quando percebemos que as diferentes mensagens, embora cada uma delas seja um E, no entanto, desempenham papéis diferentes na série de mensagens. Quando recebemos sessenta mensagens (em outras palavras, quando temos o conhecimento C_{60}), então sabemos que sessenta dos sessenta e quatro estados possíveis do universo são excluídos. Ainda restam quatro estados possíveis; isto é, nosso conhecimento C_{60} significa que o universo está em um dos quatro estados restantes. A sexagésima primeira mensagem exclui entre esses quatro possíveis um outro adicional; portanto, o intervalo daqueles que ainda estão em aberto diminui de quatro para três.

Tabela V

M	$cont_D(C_m)$	$cont_D(E_m/C_{m-1})$	$inf_D(C_m)$	$inf_D(E_m/C_{m-1})$
1	1/64	1/64	0.0228	0.0228
2	1/64	2/64	0.0459	0.0231
3	1/64	3/64	0.0693	0.0234
4	1/64	4/64	0.0931	0.0238
5	1/64	5/64	0.1174	0.0242
6	1/64	6/64	0.1421	0.0247
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
59	1/64	59/64	3.6781	0.2630
60	1/64	60/64	4.0000	0.3219
61	1/64	61/64	4.4151	0.4151
62	1/64	62/64	5.0000	0.5849
63	1/64	63/64	6.0000	1.0000
64	1/64	1	∞	∞

Pela mensagem de sessenta segundos, o alcance é reduzido de três para dois, e isso pode muito bem ser considerado como um acréscimo mais forte ao nosso conhecimento do que o diminuir de quatro para três. Neste momento, apenas duas possibilidades são deixadas em aberto. A sexagésima terceira mensagem nos dá informações sobre qual dessas duas sentenças restantes é a verdadeira e, portanto, completa o nosso conhecimento do universo. Portanto, este passo tem um grande peso, mais do que qualquer um anterior. Depois desse passo, nada pode ser adicionado ao nosso conhecimento de forma consistente. A sexagésima quarta mensagem é incompatível com a conjunção dos sessenta e três precedentes. Se esta mensagem é, no entanto, adicionada, então este é um passo ainda mais pesado que leva à contradição. A mensagem factual mais forte, que é uma descrição do estado, uma conjunção de 6 sentenças, carrega 6 unidades de informação medidas pelo . As únicas mensagens que transportam mais unidades de informação e, em seguida, por necessidade infinitamente, muitas dessas unidades, são as mensagens que contradizem a si mesmas ou mensagens anteriores.

§10. *cont** e *inf**

> Duas funções especiais $cont_I$ e inf_I, $cont^*$e inf^*, são definidas e teoremas em relação a elas desenvolvidos. Essas funções são baseadas em m_I- m^* - que atribui valores iguais a todas as descrições de estrutura, ou seja, disjunções de descrições-de-estado isomórficas. Uma vez que m^* pareça ter um especial status entre as várias funções m_I, $cont^*$e inf^* são considerados de importância especial. Vários cálculos e tabelas referentes a estas funções são apresentados.

Vamos agora definir e investigar duas funções-I especiais que podem vir a ser de especial importância. Elas são baseadas na função *m* definida em [Prob.] §110 essencialmente como uma função *m* adequada que tem os mesmos valores para todas as descrições de estrutura, isto é, disjunções de descrições-de-estado isomórficas.

Lembrando a definição de 'sentenças isomórficas' dada em §6, o leitor facilmente verá que nossos $\mathcal{L}_3^2$ tem exatamente 20 descrições de estrutura. Sendo o Z de $\mathcal{L}_3^2$ como apresentado na Tabela I; Z_1, Z_2, . . . , Z_{64}. Então as descrições de estrutura $'T_1'$, $'T_2'$, ... , $'T_{20}'$, serão:

T_1: Z_1
T_2: Z_2
T_3: Z_3
T_4: Z_4
T_5: $Z_5 \vee Z_6 \vee Z_7$
T_6: $Z_8 \vee Z_9 \vee Z_{10}$
T_7: $Z_{11} \vee Z_{12} \vee Z_{13}$
T_8: $Z_{14} \vee Z_{15} \vee Z_{16}$
T_9: $Z_{17} \vee Z_{18} \vee Z_{19}$
T_{10}: $Z_{20} \vee Z_{21} \vee Z_{22}$
T_{11}: $Z_{23} \vee Z_{24} \vee Z_{25}$
T_{12} : $Z_{26} \vee Z_{27} \vee Z_{28}$
T_{13} $Z_{29} \vee Z_{30} \vee Z_{31}$
T_{14} : $Z_{32} \vee Z_{33} \vee Z_{34}$
T_{15}: $Z_{35} \vee Z_{36} \vee Z_{37}$
T_{16}: $Z_{38} \vee Z_{39} \vee Z_{40}$
T_{17} $Z_{41} \vee Z_{42} \vee Z_{43} \vee Z_{44} \vee Z_{45} \vee Z_{46}$
T_{18} $Z_{47} \vee Z_{48} \vee Z_{49} \vee Z_{50} \vee Z_{51} \vee Z_{52}$

T_{19} $Z_{53} \vee Z_{54} \vee Z_{55} \vee Z_{56} \vee Z_{57} \vee Z_{58}$
T_{20}: $Z_{59} \vee Z_{60} \vee Z_{61} \vee Z_{62} \vee Z_{63} \vee Z_{64}$

, $m^*(Z_5) =$ '64) = 1/120.

Para todo i, m^* $(\boldsymbol{T_i}) = 1/20$, portanto $m^*(Z_1) = m^*(Z_2) = m^*(Z_3) = m^*(Z_4) =$ 1/20, $m^*(Z_5) = \ldots = m^*(Z_{40}) = 1/60$, $m^*(Z_{41}) = \ldots = m^*(Z_{64}) = 1/120$.

Em [Cont.] §18, é apresentado um argumento que mostra que a função c^* baseada em m^* é, em certo sentido, mais simples que outras funções c_P. *Explicata* para quantidade de informação com base em m^* compartilharia esse status especial.

$c^*(h,e)$, $cont^*(e)$, $cont^*(h/e)$, $inf^*(e)$ e $inf^*(h/e)$ podem ser todas expressas como simples funções de $m^*(e)$ e $m^*(e \wedge h)$:

$$c^*(h,e) = \frac{m^*(e \wedge h)}{m^*(e)} \; (D7-3). \tag{1}$$

$$cont^*(e) = 1 - m^*(e) \; (T6-3a)\,. \tag{2}$$

$$cont^*(h/e) = m^*(e) - m^*(e \wedge h) \; (T6-8a)\,. \tag{3}$$

$$inf^*(e) = -Log\, m^*(e) \; \text{(T7-5b)}. \tag{4}$$

$$inf^*(h/e) = Log\, m^*(e) - Log\, m^*(e \wedge h) \; \text{(fórmula (1) em §8)}. \tag{5}$$

Tome *e* sendo '$Ma \wedge Mb$' e *h* sendo 'Mc'. Então, por inspeção da Tabela I, vemos que

$$m^* \text{ ('Ma. Mb')} = 2 \times \frac{1}{20} + 10 \times \frac{1}{60} + 4 \times \frac{1}{1Z0} = 0.3.$$

Observe que m_D (e) = 0. 25. O maior valor de m^* é devido à relevância instancial. Também temos

$$m^* \; ('Ma \wedge Mb \wedge Mc') = 2 \times \frac{1}{20} + 6 \times \frac{1}{60} = 0.2.$$

Consequentemente

c^* $('Mc', 'Ma \wedge Mb') = \frac{0.2}{0.3} = \frac{2}{3}$.

$cont^*('Ma \wedge Mb') = 0.7$.

$cont^*('Mc'/'Ma \wedge Mb') = 0.3 - 0.2 = 0.1$;

Por outro lado,

$cont_D\ ('Mc'/\,'Ma \wedge Mb') = 0.125.$
$inf^*\ ('Ma \wedge Mb') = -\,Log\ 0.3 = Log\ 10 - Log\ 3 = 1.\ 7370,$

contra um valor inf_D de 2. Finalmente,

$inf^*\ ('Mc'/\,'Ma \wedge Mb') = \text{Log}\ \frac{0..3}{0.2} = \text{Log}\ 3 - \text{Log}\ 2 = 0.\ 5849$

enquanto que o valor relativo inf_D é 1.

Talvez valha a pena investigar agora outra linguagem de amostra, desta vez com apenas um predicado primitivo e *n* constantes individuais distintas. Neste caso, c^* produz os mesmos valores que a regra de sucessão de Laplace. (Veja [Prob.] §110E.) Seja *e* uma conjunção de $s < n$ sentenças básicas com *s* constantes individuais distintas, entre elas sentenças atômicas com negações 'P' e *s*-s_1 de tais. Deixemos *h* ser ' ' onde 'b' é uma constante individual que não ocorre em *e*. Então, o seguinte é válido (de acordo com a fórmula (4), [Prob.] P. 566, *cf*. observação a D6-1*i*).

$$m^*(e) = \frac{s_1!\,(s-s_1)!}{(s+1)!} = \frac{1}{(s+1)\binom{s}{s_1}}\,. \qquad (6)[7]$$

$$m^*(e \wedge h) = \frac{(s_1+1)!\,(s-s_1)!}{(s+2)!} = m^*(e) \times \frac{s_1+1}{s+2}. \qquad (7)$$

$$c^*(h,e) = \frac{s_1+1}{s+2}\ ((1)\ ,(6)\ ,(7))\ . \qquad (8)$$

$$cont^*(h/e) = m^*(e) \times (1 - \frac{s_1+1}{s+2})((3), (6)\ (7)) = m^*(e) \times \frac{s-s_1+1}{s+2}. \qquad (9)$$

Para ter um exemplo numérico, assumimos s=10. Nós temos

$$m^*(e) = \frac{1}{11\binom{10}{s_1}}. \qquad (10)$$

$$m^*(e \wedge\ h) = m^*(e) \times \frac{s_1+1}{12}. \qquad (11)$$

$$c^*(h,e) = \frac{s_1+1}{12}. \qquad (12)$$

$$cont^*(h/e) = m^*(e) \times \frac{11-s_1}{12}. \qquad (13)$$

[7] No texto original, a fórmula está parcialmente apagada. Podendo ser vista apenas como $m^*(e) = \frac{s_1!\,(s-s_1)!}{} = \frac{1}{(s+1)\binom{s}{s_1}}$. O restante da equação é uma inferência feita a partir das equações anteriores e da referência à "[Prob.] p.566". (N.T.)

$$\inf^*(\mathrm{e}) = \mathrm{Log}\frac{1}{m^*(e)}. \tag{14}$$

$$\inf^*(h/e) = \mathrm{Log}\,\frac{12}{s_1+1}. \tag{15}$$

Os valores dados na Tabela VI são calculados de acordo com estas fórmulas.

Tabela VI

s_1	$m^*(e)$	$m^*(e \wedge h)$	$c^*(h,e)$	$cont^*(e)$	$cont^*(h/e)$	$\inf^*(e)$	$\inf^*(h/e)$
0	0.09091	0.0076	0.0833	0.90909	0.08333	3.459	3.585
1	0.00909	0.0015	0.1667	0.99091	0.00758	6.781	2.585
2	0.00202	0.0005	0.2500	0.99798	0.00152	8.751	2.000
3	0.00076	0.0003	0.3333	0.99924	0.00051	10.366	1.585
4	0.00043	0.0002	0.4167	0.99957	0.00025	11.174	1.263
5	0.00036	0.0002	0.5000	0.99964	0.00018	11.437	1.000
6	0.00043	0.0003	0.5833	0.99957	0.00018	11.174	0.778
7	0.00076	0.0005	0.6666	0.99924	0.00025	10.366	0.585
8	0.00202	0.0015	0.7500	0.99798	0.00051	8.751	0.415
9	0.00909	0.0076	0.8333	0.99091	0.00152	6.781	0.263
10	0.09091	0.0833	0.9167	0.90909	0.00758	3.459	0.126

Além dessas fórmulas, temos, claro, também

$$m^*(h) = \frac{1}{2}. \tag{16}$$

$$cont^*(h) = \frac{1}{2}. \tag{17}$$

$$inf^*(\mathrm{h}) = 1. \tag{18}$$

Alguns comentários sobre a Tabela VI podem ser indicados. As colunas para $m(e)$ e $m^*(e \wedge h)$ mostram que esta função m_I, como é de se esperar de qualquer função m_I adequada, valoriza a homogeneidade; isto é, aqueles estados nos quais a diferença absoluta entre os indivíduos que têm P e os que não têm P são maiores, são tratados como inicialmente mais prováveis. Quando a evidência afirma que 5 indivíduos têm P e 5 outros não têm P, a última coluna mostra que o inf^*-valor da nossa hipótese, a qual afirma que um décimo primeiro indivíduo tem P, é apenas 1. Por isso, é o mesmo que o inf^*-valor absoluto desta hipótese. Quanto maior o número de indivíduos com P, de acordo com as evidências, maior o c^* (h, e) e o menor $inf^*(h/e)$. $cont^*(h/e)$ no entanto, se comporta de maneira diferente. Atinge seu mínimo para valores intermediários de **F**, mas aumenta tanto quanto s_1 aumenta de 6 para 10 e quando diminui de 5 para 0.

§ 11. Estimativas da quantidade de informação

Um cientista está frequentemente interessado no valor expectativa da quantidade de informações transmitidas pelo resultado de um experimento a ser realizado. Se os vários resultados possíveis podem ser expressos por $h_1, h_2, \ldots, h_n$, tais que essas sentenças sejam exclusivas dos pares e sua disjunção L-verdadeira sobre a evidência dada e, em resumo, quando H = $\{h_1, h_2, \ldots, h_n\}$ é um sistema exaustivo em *e*, a estimativa da quantidade de informação transportada por H em relação a *e* é dada pela fórmula

$$est\ (in, H, e)\ = \sum_{p=1}^{n} c(h_p, e) \times\ in(h_p/e)$$

Várias fórmulas para $est(cont, H, e)$, $est\ (in, H, e)$ e outras funções baseadas sobre elas são derivadas. Os conceitos de estimativa posterior da quantidade de informação, quantidade de especificação, estimativa da estimativa posterior e estimativa da quantidade de especificação são definidos, e vários teoremas sobre eles provados. Uma simples aplicação ilustrativa é dada.

Se uma experiência é realizada, os resultados possíveis dos quais são expressos em *n* sentenças $h_1, h_2, \ldots, h_n$ (ou em *n* frases L-equivalentes a elas), podemos computar a quantidade de informação que cada resultado possível transmitiria, assumindo que uma função *m* foi definida para todas as sentenças da linguagem em que os *h's* são formulados. Conquanto o resultado real não for conhecido, a quantidade de informação que carrega também é desconhecida. Mas, para certos propósitos, é importante ter uma boa estimativa deste montante. A situação é análoga à existente muito frequentemente em investigações, onde uma certa magnitude é desconhecida e se tem que trabalhar em vez disso com uma estimativa dessa magnitude.

Para dar um exemplo bruto, mas suficientemente ilustrativo: Imagine um termômetro que é dividido de maneira não convencional em três regiões, de modo que na região 1 o ponteiro indica Quente, na região 2 Temperado, e na região 3 Frio. Deixe o termômetro ser lido em um lugar onde, de acordo com as evidências disponíveis, a maioria das leituras anteriores indicava algumas Temperado, e apenas algumas poucas Quente. Desde a mesma distribuição (aproximadamente) espera-se que, para futuras leituras, uma medida adequada de informação seja atribuída a frase 'Frio (t_1)' (onde t_1 é um ponto no tempo no futuro, isto é, um não mencionado em evidência), um valor menor, relativo à evidência, do que para 'Temperado (t_1)' que novamente terá um valor menor que 'Quente (t_1)'. Deixe essas sentenças serem h_1, h_2 e h_3, respectivamente. O que seria uma estimativa razoável da quantidade de informação de um futuro que se espera que a observação leve? Pode-se, a princípio, pensar em tomar a média aritmética das três quantidades de informação, isto é,

$$\frac{in(h_1/e) + in(h_2/e) + in(h_3/e)}{3},$$

mas uma pequena reflexão mostrará que isso seria totalmente inadequado. Os valores têm que ser ponderados de forma diferente. Parece bastante natural ter pesos apropriados aqui, bem como, em geral, os graus de confirmação que as sentenças h_1, h_2 e h_3 têm na evidência disponível. (Para uma discussão mais completa deste procedimento, veja [Prob.] Cap. ix.) Chegamos, portanto, ao valor

$$c\ (h_1, e) \times in(h_1/e) + c(h_2, e) \times in(h_2/e) + c(h_3, e) \times in(h_3/e)$$

ou, na abreviatura habitual conveniente,

$$\sum_{p=1}^{3} c(h_p, e) \times\ in(h_p/e)$$

Expressões deste tipo são bem conhecidas na teoria da probabilidade e estatística (com a subfórmula de grau de confirmação geralmente substituída por uma fórmula de frequência relativa) sob o nome 'a expectativa matemática (ou esperança) de . . .' , no nosso caso, '... da quantidade de informação transportada pela observação a ser feita em t_1'.

Em geral, sempre que temos uma classe de sentenças H = {h_1..., h_n} tal que a evidência disponível e L-implica $h_1 \vee h_2 \vee ... \vee h_n$ bem como ~ $(h_i \wedge h_j)$, para todos $i \neq j$, nós devemos dizer que H é um sistema exaustivo relativo para *e*, e a expressão

$$\sum_{p=1}^{n} c(h_p, e) \times\ in(h_p/e)$$

será chamada 'a estimativa (c-média) da quantidade de informação transportada por membros de) H em relação a *e* ', simbolizada por' est (in, H, *e*) '.

Até agora, nossa discussão tem se processado em um nível parcialmente pré-sistemático, e parcialmente sistemático. Para mudar para um tratamento completamente sistemático, obviamente só temos que substituir o *explicandum* 'in' por uma ou outra de suas *explicata*. Nós definimos:

D11-1. Seja H, h_p, *e* como acima. Então

$$est\ (cont, H, e)\ =_{DF}\ \sum_p c(h_p, e) \times cont(h_p/e)$$

D11-2. Seja H, F, e como acima. Então

$$est\ (inf, H, e)\ =_{DF}\ \sum_p c(h_p, e) \times inf(h_p/e)$$

E (exemplo) 11-1 Vamos, por exemplo, com respeito aos nossos $\mathcal{L}_3^2$, h_1= 'M*c*', h_2 = 'F*c*', H = $\{h_1, h_2\}$, *e* = $'Ma \wedge\ Mb'$. Com base na Tabela I, algumas fórmulas na seção anterior, e as duas fórmulas seguintes que o leitor facilmente será capaz de derivar para si mesmo, nomeadamente,

$$cont^*\ ('Mc'/'Ma \wedge\ Mb')\ =\ 0.1$$

e

$$cont^*\ ('Fc'/'Ma \wedge\ Mb')\ =\ 0.2$$

obtemos agora

est $(cont^*, H, e) = \frac{2}{3} \times 0.1 + \frac{1}{3} \times 0.2 = 0.133$

e

$est(inf^*, H, e) = -(\frac{2}{3} \times\ \mathrm{Log}\ \frac{2}{3}\ + \frac{1}{3} \times\ \mathrm{Log}\ \frac{1}{3}) = 0.918.$

$est(inf_D, H, e)$, por outro lado, é igual a 1, é claro.

E11-2. Seja h_1, h_2 e H como antes, mas temos agora

e = '$Ma \wedge Mb \wedge Ya \wedge Yb \wedge Yc$'.

Então

$$m^*(e) = \frac{1}{20} + \frac{1}{60} = \frac{1}{15},$$
$$m^*(\mathrm{e} \wedge h_1)\ =\ \frac{1}{20},$$

$$m^*(e \wedge h_2) = \frac{1}{60}.$$

Consequentemente

$cont^*(h_1/e) = \frac{1}{60}$,
$cont^*(h_2/e) = \frac{1}{20}$,
$\inf^*(h_1/e) = 0.4151$,

$$\inf^*(h_2/e) = 2,$$

$$c^*(h_1, e) = \frac{3}{4},$$

e

$$c^*(h_2, e) = \frac{1}{4}.$$

Consequentemente

$est(cont^*, H, e) = \frac{3}{4} \times \frac{1}{60} + \frac{1}{4} \times \frac{1}{20} = \frac{1}{40}$,

e

est(inf^*, H, e) $= \frac{3}{4} \times 0.4151 + \frac{1}{4} \times 2 = 0.811$

(enquanto est (inf_D, H, e) é igual a 1).

Para os teoremas a seguir, é sempre assumido que H, h_p e e preencham as condições acima mencionadas.

T11- 1

est(cont, H, e) $= \sum_p \frac{m(h_p.e)}{m(e)} \times m(e) \times (1 - c(h_p, e)) \quad (T7-14)$

$$= \sum_p m(h_p \wedge e) \times (1 - c(h_p, e))$$

$$= \sum m(h_p \wedge e) \times c(\sim h_p, e)$$

$$= \frac{1}{m(e)} \sum m(h_p \wedge e) \times m(\sim h_p \wedge e)$$

$$= \frac{1}{m(e)} \sum m(h_p \wedge e) \times (1 - m(h_p \wedge e))$$

$$= \sum c(h_p, e) \times m(\sim h_p \wedge e)$$

$$= c\ (e,t) \sum c(h_p,e)\ \times c(\sim h_p,e)$$

Seja K = $\{k_1,\ldots,k_n\}$ um sistema exaustivo em relação a e. Então de bem conhecidos. teoremas na teoria das desigualdades, o seguinte teorema pode ser derivado:

Tll-2. Seja $(k_i,e) = c(k_j,e)$ para todos i e j (portanto = l / n), e que haja pelo menos um par i e j tal que $c(h_i,e) \neq c(h_j$, e). Então

$est(cont, K, e) > \ est(cont, H, e)$.

T11-3. Para n fixo, est (cont, H_i, e) é um máximo para aqueles todos H_i cujos membros tem os mesmos c-valores em e. Consequentemente

$$\max_i[est(cont, H_i, e)]\ =\ m(e)\ \times \frac{n-1}{n}.$$

(Isto é, naturalmente, também o valor contido de cada h_p^i pertencente a estes H_i.)

T11-4. Para n fixo, est (cont, H_i, e) é um mínimo para aqueles H_i um membro dos Quais tem o c-valor 1 em e (e, portanto, todos os outros membros, o c-valor 0 em e);

consequentemente

$$\min_i[est(cont, H_i, e)]\ = 0$$

Teoremas semelhantes a T2, T3 e T4 podem ser obtidos para o segundo *explicatum* inf.

Vamos primeiro declarar uma transformação de D2, de acordo com T7-13:

T11-5.

$$\text{est}(\text{inf}, H, e) = \sum\ c(h_p,e) \times \text{Log}\ \frac{1}{c(h_p,e)}$$
$$= \sum\ c\ (h_p,e)\ \times\ Log\ c(h_p,e)$$

Temos agora

T11-6 Seja $c(k_i,e) = c(k_j$,e) para todos os i e j (portanto = l / n), e que haja pelo menos um par i e j tal que $c(h_i,e) \neq c(h_j,e)$. Então

$$est(inf, K, e) > \ est(inf, H, e).$$

Tll-7. Para n fixado, est (inf, H_i, e) é um máximo para aqueles cujos todos os membros tenham os mesmos valores c em e; consequentemente

$\max_i[est(\inf, H_i, e)] =$ Log n.

(Isto é, naturalmente, também o valor de cada h_p^i pertencente a estes H_i)

Tll-8. Para n fixado, $est(\inf, H_i, e)$ é um mínimo para aqueles H_i um membro do qual tem o valor c 1 em e (e, portanto, todos os outros membros têm o valor c 0 em e).

Consequentemente

$\min_i[est(\inf, H_i, e)] = 0$

Uma expressão análoga a

$$' - \sum c(h_p, e) \times \operatorname{Log} c(h_p, e)'$$

mas com grau de confirmação substituído por probabilidade (estatística), desempenha um papel central na teoria da comunicação, bem como em certas formulações de mecânica estatística, onde a probabilidade em questão é aquela de um sistema que está na célula p do seu espaço fásico. Na mecânica estatística, na formulação dada por Boltzmann, esta expressão é dita para medir a entropia do sistema. Em analogia a isso, alguns teóricos da comunicação chamam a expressão correspondente, que surge quando as probabilidades em causa são as frequências relativas (esperadas) da ocorrência de determinadas mensagens, a entropia deste sistema de mensagens. Outros termos, usados como sinônimos, embora infelizmente sem qualquer esforço real para esclarecimento terminológico, eram incerteza, escolha e, até mesmo simplesmente, confusamente, informação.

Seja H e K sistemas exaustivos em relação a e, seja H contendo n membros, e K contendo m membros. Seja $'H \wedge K'$ abreviação para $'\{h_1 \wedge k_1, h_1 \wedge k_2, \ldots, h_1 \wedge k_m, h_2 \wedge k_1, \ldots, h_n \wedge k_m\}'$. Então definimos

D11-3.

$$est(in, H \wedge K, e) =_{Df} \sum_{q=1}^{m} \sum_{p=1}^{n} c(h_p . k_q, e) \times in(h_p . k_q / e)$$

Com relação ao *explicatum* inf, o seguinte teorema pode ser provado:

T11-9. $est(in, H \wedge K, e) \leq est(inf, H, e) + est(inf, K, e)$, onde a igualdade vale apenas se, para todos p e q, $c(h_p \wedge k_q, e) = c(h_p, e) \times c(k_q, e)$, em outras palavras, quando os h's e os k's são indutivamente independentes de e (em relação a essa função-m em que c é sediada).

E11-3. Seja $e = 'Ma \wedge Mb \wedge Ya \wedge Yb'$, $h_1 = 'Mc'$, $h_2 = 'Fc'$, $k_1 = 'Yc'$, $k_2 = 'Oc'$, $H = \{h_1, h_2\}$ e $K = \{k_1, k_2\}$

Então,

$$H \wedge K = \{h_1 \wedge k_1, h_1 \wedge k_2, h_2 \wedge k_1, h_2 \wedge k_2\}.$$

Nós temos

$$m^*(e) = \frac{1}{10},$$
$$m^*(h_1 . k_1 . e) = \frac{1}{20},$$
$$m^*(h_1 \wedge k_2 \wedge e) = m(h_2 \wedge k_1 \wedge e) * = m^*(h_2 \wedge k_2 \wedge e) = \frac{1}{60},$$
$$c^*(h_1 \wedge k_1, e) = \frac{1}{2},$$
$$c^*(h_2 \wedge k_2, e) = c^*(h_2 \wedge k_1, e) = c^*(h_2 \wedge k_2, e) = \frac{1}{6},$$
$$cont^*(h_1 \wedge k_1 / e) = \frac{1}{20},$$
$$cont^*(h_1 \wedge k_1 / e) = ... = \frac{1}{12},$$
$$inf^*(h_1 \wedge k_1 / e) = 1,$$
$$inf^*(h_1 \wedge k_Z / e) = ... = 2.585.$$

Consequentemente

est($cont^*$, H$\wedge$ K, e) = $\frac{1}{15}$

e

est(inf^* , H$\wedge$ K, e) = 1.792

est(inf^* , H, e) = est(inf^* , K, e) = 0.918

Verificamos que

$$est\ (inf^*, H \wedge K, e) < est\ (inf^*, H, e) + est\ (inf^*, K, e),$$

os h's e os k's não estão sendo indutivamente independentes neste e em relação a m *. Eles são, no entanto, independentes em relação a, e de fato

$$est\ (inf_D, H \wedge K, e) = 2 = est\ (inf_D, H, e) + est\ (inf_D, K, e).$$

Em geral, $est(in, H, e)$ será diferente de $est(in, H, e.K)$, onde k é uma sentença que foi adicionada à evidência anterior e. Uma vez que '$est\ (in, H, e.k)$' e expressões similares são de grande importância, vale a pena dar-lhes um nome especial. Vamos chamar a estimativa posterior (da quantidade de informação transportada por H na evidência composta por e e k). A expressão '$est\ (in, H, e)$' será então chamada, para maior clareza, de a estimativa prévia (de...). Muitas vezes, é importante investigar como tal prévia estimativa foi alterada por meio de algumas evidências adicionais. Vamos, portanto, dar também à diferença entre a estimativa anterior e posterior, um nome especial, a quantidade de especificação de H através de k em e, e denotar esta função por um símbolo especial '*sp* (in, H, *k*, *e*)':

D11- 4. $sp\ (in, H, k, e) =_{DF} est(in, H, e) - est(in, H, e.k)$.

E11-4. Seja e, H e k_1 como em E3. Então $e \wedge k_1$ é o e de E2. Assim sendo $est(inf^*, H, e \wedge k_1) = 0{,}811$. Como $est\ (inf^*, H, e) = 0.918$ (de E3), temos $sp(inf^*, H, k_1, e) = 0{,}918 - 0{,}811 = 0,107$.

Pode ser facilmente visto que *sp* (in, H, *k*, *e*) = 0 se (mas não somente se) k é indutivamente independente dos *h's* em *e*. Caso contrário, *sp* pode ser positivo ou negativo. Seu valor máximo é obviamente igual a est (in, H, *e*) em si. Este valor será obtido quando $e \wedge k$ L-implica um dos *h*'s. Neste caso, H é maximamente especificado através de *k* em *e*.

Muitas vezes surgem situações em que o evento indicado em k ainda não ocorreu ou, pelo menos, em que não sabemos se ocorreu ou não, mas sabemos apenas que ele ou algum outro evento pertencente a um sistema exaustivo de eventos ocorrerá ou terá ocorrido. Em tais circunstâncias, faz sentido pedir o valor esperado da estimativa posterior da quantidade de informação transportada por H em e e (algum membro do sistema exaustivo) K. Somos levados à estimativa (*c*-média) desta estimativa posterior que denotaremos por '$est(in, H/K, e)$' e definiremos como

Dl1-5.

$$est(in, H/K, e) =_{DF} \sum_{q=1}^{m} c(k_q, e) \times est(in, H, e \wedge k_q).$$

E11-5. Seja e, H e K como em E3. Então

$$est(in, H/K, e) = \tfrac{2}{3} \times 0.811 + \tfrac{1}{3} \times 1 = 0.874.$$

A notação de traço foi escolhida em vez de uma notação de vírgula mais neutra, porque o teorema a seguir, que está em certa analogia com as definições de quantidades relativas de informação, é válido.

Tll -10. est (in, H / K, e) = est (in, H. K, e) - est (in, K, e).

Prova:

$$\begin{aligned} est(in, H/K, e) &= \sum_q c(k_q, e) \sum_p c(h_p, e \wedge k_q) \times in(h_p/e \wedge k_q) \\ &= \sum_q \sum_p c(k_q, e) \times c(h_p, e.k_q) \times in(h_p/e.k_q) \\ &= \sum_q \sum_p c(h_p \wedge k_q, e) \times in(h_p/e \wedge k_q) \\ &= \sum_q \sum_p c(h_p \wedge k_q, e) x\, [in(h_p \wedge k_q/e) - in(k_q/e)] \\ &= est(in, H \wedge K, e) - \sum_q c(k_q, e) \times in(k_q/e) \\ &= est(in, H.K, e) - est(in, K, e). \end{aligned}$$

De fato, $est(inf^*, H \wedge K, e) - est(inf^*, K, e) = 1.792$ (de E3) – 0.918 (de E3) = 0.874 (como em E5)

Frequentemente, alguém estará interessado em uma estimativa da quantidade de especificação de H em *e* através de K. Esta função será simbolizada por '*sp* (in, H, K, *e*)'. Sua definição é

D11-6.

$$sp(in\,, H, K, e) =_{Df} \sum_q c\big(k_q, e\big) \times sp\big(in\,, H, k_q, e\big).$$

Ell-6. Seja e, H e K como em E3. Então

$$sp(\mathrm{inf}^*, H, K, e) = \tfrac{2}{3} \times 0.107 \text{ (from E4)} + \tfrac{1}{3} \times (-0.082) \text{ (computado do mesmo modo)} = 0.044.$$

Vemos imediatamente que o seguinte teorema é válido:

T11 -11. $sp(in, H, K, e) = est(in, H, e) - est(in, H/K, e)$.

De fato, $est\,(inf^*, H, e)\ - est\,(inf^*, H/K, e)\ =\ 0.918\,(E3)\ -\ 0.874\,(E5)\ =$ 0.044 (como em E6).

Embora possa acontecer que, para alguns *q*, *sp* (in, H, **F**, *e*) seja negativo, pode ser provado que *sp* (in, H, K, *e)* nunca é negativo, em outras palavras, que a estimativa posterior é no máximo igual à estimativa da anterior.

T11-12. $sp(in, H, K, e) \geq 0$, com igualdade mantida sse os h's e os k's são indutivamente independentes.

Combinando T10 e T11, conseguimos

T11-13. $sp(in, H, K, e) = est(in, H, e) + est(in, K, e) - est(in, H \wedge K, e)$.

De T13 segue imediatamente o seguinte teorema da simetria ou mutualidade de especificação:

T-14. $sp(in, H, K, e) = sp(in, K, H, e)$.

Para ilustrar a importância e uso das funções definidas nesta seção, vamos elaborar um exemplo numérico diferente, ainda que artificialmente simplificado, para facilitar a computação. Seja h_1 'Jones é brilhante', h_2 'Jones é mediano (em inteligência)', e h_3 'Jones é maçante'. Alguém que está interessado na inteligência de Jones faz ele passar por um certo teste. Deixe agora k_1 ser 'Jones atinge mais de 80 por cento (neste teste) ', k_2 'Jones atinge entre 60% e 80% ', e k_3 ' Jones alcança menos de 60% '. Sejam os seguintes graus de confirmação que mantenham a evidência disponível, de acordo com alguma função-*m*:

$$c\,(h_1, e) = c(h_3, e) = \tfrac{1}{4}$$

$$c(h_2, e)\ =\ \frac{1}{2}$$

$$c\,(k_1, e \wedge h_1)\ =\ c(k_2, e \wedge h_1)\ =\ c(k_2, e \wedge h_2)\ =\ c(k_2, e \wedge h_3)\ =\ c(k_3, e \wedge h_3)\ = \frac{1}{2}$$

$$c\,(k_1, e \wedge h_2)\ =\ c(k_3, e \wedge h_2)\ =\ \frac{1}{4}.$$

(*All other* $c(k_q, e \wedge h_p) = 0$.) A Figura 2 pode ajudar a visualizar a situação.

Para os seguintes cálculos, o *explicatum* inf será usado. Primeiro nós computamos com a ajuda de T5 o valor de $est(inf, H, e)$ em nosso exemplo.

$est(inf, H, e) = \frac{1}{4}$ Log 4 $+ \frac{1}{2}$ Log 2 $+ \frac{1}{4}$ Log 4 $= 1.5$.

Para avaliar $est(inf, K, e)$ temos primeiro que encontrar os vários $c(k_q, e)$. Estes podem ser facilmente lidos no diagrama.

$$c\ (k_1, e) = c(k_3, e) = \frac{1}{4},$$
$$c(k_2, e)\ = \frac{1}{2}.$$

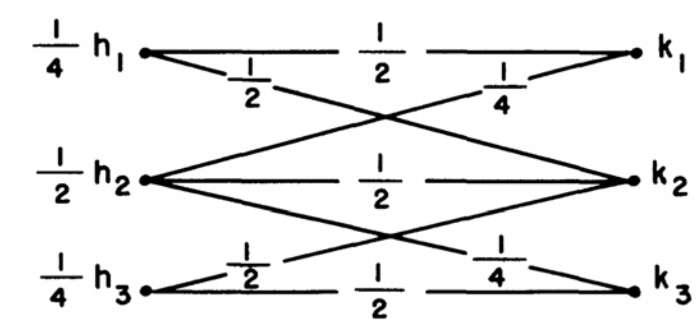

Fig. 2

Uma vez que $c(k_i, e) = c(h_i, e)$ para todo i (isso é pura coincidência), nós temos

$$est(inf, K, e) = 1{,}5.$$

Para $est(inf, H \wedge K, e)$ obtemos, novamente por simples inspeção do diagrama,

$$est(inf, H \wedge K, e)\ =\ 6\ \times \frac{1}{8}\ Log\, 8\ +\ 1 \times \frac{1}{4}\, Log\, 4\ =\ 2.75.$$

Isso verifica T9. É óbvio que nem todos os h e k são indutivamente independentes.

Para encontrar os vários $est\ (inf, H, e \wedge k_q)$, computamos primeiro todos os $c(h_p, e \wedge k_q)$.

Nós temos

$$c\ (h_1, e \wedge k_1) =\ c(h_2, e \wedge k_1) =\ c(h_2, e \wedge k_2) =\ c(h_2, e \wedge k_3) =\ c(h_3, e \wedge k_3) = \frac{1}{2}.$$
$$c(h_1, e \wedge k_2)\ =\ c(h_3, e \wedge k_2)\ \frac{1}{4}.$$

(Todos os outros $c(h_p, e \wedge k_q) = 0$.) Portanto, temos

$$est(inf, H, e \wedge k_1) = est(inf , H, e \wedge k_3) = 1,$$
$$est(inf, H, e \wedge k_2) = 1.5$$

Por isso, temos, de acordo com D4,

$$sp(inf, H, k_1, e) = sp(inf, H, k_3, e) = \frac{1}{2}.$$
$$sp(inf, H, k_2, e) = 0.$$

O último resultado é de especial importância. E, de fato, se Jones alcança entre 60 por cento e 80 por cento em seu teste, nós somos "tão inteligentes quanto antes", sabemos exatamente como tanto quanto sabíamos antes. A adição de k_2 às nossas evidências deixou os valores-c dos h's inalterados, k_1 é indutivamente irrelevante para os h's, e nosso conhecimento não se tornou mais específico através desta adição. A situação é diferente em relação aos dois outros resultados do teste. Nos outros casos, nosso conhecimento tornou-se mais específico. Isto aparece mesmo no nível qualitativo: Antes do teste, Jones poderia ter sido brilhante, médio ou sem graça. Após o teste, sabemos que ele não é chato se o resultado for k_1, e que ele não é brilhante, se o resultado for k_3. Mas é preciso ter cuidado com esse argumento. Uma redução do número de possibilidades nem sempre implica um aumento de especificidade da situação. Se a distribuição de probabilidade das possibilidades restantes for muito mais uniformemente espalhada do que a das possibilidades iniciais, a situação pode tornar-se em um certo importante sentido, menos específico. Exemplos podem ser facilmente construídos. No nosso caso, no entanto, há um aumento real na especificidade, embora não seja grande.

Parece razoável medir um aspecto da eficácia deste teste de inteligência pela estimativa da quantidade de especificação. Pode-se comparar a eficácia de vários testes propostos dessa maneira. No nosso caso, de acordo com o D6,

$$sp(inf, H, K, e) = \frac{1}{4} \times \frac{1}{2} + \frac{1}{2} \times 0 + \frac{1}{4} \times \frac{1}{2} = \frac{1}{4},$$

um resultado que poderia, claro, também ter sido obtido a partir de T13. Aliás, segue que o teste foi muito pobre. Considerando uma medida direta da inteligência de Jones, se fosse apenas possível, seria de esperar que gerasse 1,5 unidades de informação, o teste mencionado pode ser esperado para nos dar apenas 0,25 de uma unidade de informação sobre a Inteligência de Jones. A diferença entre 1. 5 e 0. 25, ou seja, 1. 25, é o valor de $est(inf, H/K, e)$, de acordo com o Tll. O mesmo valor seria obtido usando D5 ou T10. Podemos dizer que aplicando o teste em vez de medir a inteligência diretamente, devemos nos contentar com a expectativa de uma "perda" de 1. 25 unidades de informação. O correlato dessa função na teoria da comunicação tem sido chamado por Shannon (1948, p. 36)* de o equívoco. Com H fixo, esse teste é mais eficiente, cujo K (a classe de possíveis resultados) produz o maior valor para a estimativa da quantidade de especificação de H sobre *e* através de K, ou o mais baixo para a estimativa posterior da quantidade de informações transportadas por H sobre *e* e K.

§12. Ruído semântico, eficiência de um quadro conceitual

> Dois usos de 'ruído semântico' são distinguidos e um conceito mais geral de distorção através do ruído definido. Eficiência do quadro conceptual de uma linguagem é introduzida, tanto em relação a alguma evidência dada e de modo absoluto. O tratamento simétrico de um predicado e sua negação maximiza a eficiência inicial. Com evidências crescentes, a eficiência de uma linguagem geralmente diminui.

Sempre que um destinatário de uma mensagem não conseguir reconstruir imediatamente a mensagem como originalmente enviada, o engenheiro de comunicação descreve a situação dizendo que a mensagem foi distorcida pelo ruído. Combater o ruído é uma das suas principais tarefas. Às vezes, o receptor de uma mensagem, apesar de uma recepção que é fisicamente livre de distorção, reage de uma forma diferente da esperada pelo remetente. Tentativas foram feitas para formular esta situação em termos de ruído semântico. De fato, a mesma sentença (mais exatamente, dois tokens do mesmo tipo de sentença) pode transmitir informações diferentes (com quantidades diferentes ou iguais de informação) a duas pessoas (por exemplo, o remetente e o destinatário de uma mensagem) e isso em pelo menos dois modos: primeiro, os dois tokens, que são fisicamente iguais, são interpretados como pertencentes, a diferentes linguagens[8], e segundo, provavelmente mais comum e interessante, a informação realizada por eles é avaliada com relação a diferentes evidências. Mal-entendidos podem ser devido à confusão física ou à desvalorização semântica (ou ambos).

Além dos dois usos metafóricos do 'ruído' mencionados acima, que parecem bastante simples e não devem causar confusão se forem devidamente distinguidos entre eles mesmos e do ruído do engenheiro, parece natural usar

[8] Veja HOCKETT (1952).

esse termo também nas seguintes situações gerais. Sempre que alguém está interessado em saber se um certo evento fora de um sistema exaustivo de eventos, H, aconteceu (ou vai acontecer), mas é incapaz, por alguma razão, de observar diretamente a ocorrência desses eventos e tem que se contentar com a observação de algum evento de outro sistema exaustivo, K, onde nem todos os k_q são irrelevantes para o h_p sobre e (*cf.* [Prob.] § 65), pode-se considerar K como uma distorção ou uma transformação através do ruído de H.

Seguindo esse uso, podemos não apenas dizer que o sistema de sons que sai de um receptor de telefone é uma distorção através do ruído dos sistemas de sons que saem da boca do alto-falante e que o sistema de impressões de símbolo na saída de um teleimpressor é uma distorção do sistema de impressões de símbolos na entrada, mas também que o sistema de posições de um termômetro em um determinado momento é uma distorção do sistema das situações de temperatura naquele momento (para alguém que está interessado nas temperaturas), que o sistema de previsões meteorológicas de um determinado departamento meteorológico é uma distorção do sistema de situações meteorológicas nos momentos em que as previsões são feitas (para alguém que está interessado no tempo), e que o sistema de resultados de testes de QI é uma distorção do sistema de características de inteligência (para alguém interessado nessas características).

Se vale a pena, nos três últimos exemplos e em casos semelhantes, falar sobre a natureza se comunicando conosco e sobre nossa recepção das mensagens da natureza em um modo distorcido pelo ruído, a fim de levar para casa uma analogia útil, é questionável. Algum valor heurístico para tal forma de discurso dificilmente pode ser negado, mas a tensão que tal uso colocaria em termos como 'comunicação' ou 'mensagem' poderia ser muito alta.

Os conceitos gêmeos de "*code-efficiency*" e "*code-redundancy*" desempenham um papel importante na teoria da comunicação. Não discutiremos aqui as definições dadas a esses conceitos nem nos debruçaremos sobre suas várias aplicações (e aplicações erradas), mas daremos definições para certos correlatos semânticos que parecem ter alguma importância.

Pela eficiência do quadro conceitual (quadro referencial teórico) da linguagem L_1, com respeito a (a função de quantidade de informação) em e (a evidência) e, em símbolos: $ef\ (L_1, in, e)$, entendemos a razão de $est(in, \mathrm{H}_1, e)$, onde H_1 é a classe das sentenças completas de todos os Q-predicadores (§2) com um argumento não mencionado em e, para $\max_i\ [est(in, \mathrm{H}_1, e)]$, onde são as classes correspondentes em outras linguagens L_i cobrindo, intuitivamente falando, o mesmo terreno.[9] Parece-nos que a escolha da classe das Q-sentenças como a classe em relação à qual a eficiência de uma linguagem definida seja natural, embora certamente não seja a única plausível. A eficiência de uma linguagem, como definido aqui, muda, em função de e, com uma mudança de evidência levada em conta. Uma linguagem pode tornar-se, em certo sentido, mais ou menos eficiente com uma mudança de experiência.

> Para um habitante de Nova York, uma linguagem com os predicados 'W', 'T' e 'C', designando Quente (acima de 75 ° F), Temperado (entre 40 ° e 75 °), e Frio (abaixo de 40 °), respectivamente, seria bastante eficiente. Ele deveria se mudar para San Francisco, no entanto, sua eficiência seria altamente reduzida, porque 'T' ocorre aqui com muito mais frequência do que os outros dois.

Gostaríamos, portanto, de ter também um conceito de eficiência independente da experiência. Tal conceito é, com certeza, prontamente disponível. Temos apenas que considerar a eficiência relativa à evidência tautológica, isto é, $ef(L_1, in, t)$. Vamos chamar este conceito de eficiência inicial e denotá-lo também por ' $ef_t(L_1, in)$'. Uma linguagem em conformidade tem uma eficiência inicial máxima se, e somente se, cada uma das Q-sentenças sejam inicialmente equiprováveis, isto é, se e somente se a função-m em que é baseada, atribua valores iguais a todas as Q-sentenças com o mesmo argumento, no qual será o caso quando (mas não somente quando) esta função m trate cada primitivo predicado e sua negação em um par, como, por exemplo, m_D e todos m_I.

O tratamento simétrico de um predicado e sua negação perde um pouco a arbitrariedade com a qual tem sido frequentemente atribuída; Acontece que este tratamento, baseado psicologicamente em algum princípio de indiferença e metodologicamente sobre considerações de simplicidade, maximiza a eficiência inicial da linguagem.

Com o aumento da experiência e o estabelecimento de leis empíricas, que em sua forma mais simples é equivalente à afirmação de que certas Q-propriedades estão vazias ([Prob.] §38), a eficiência da respectiva linguagem geralmente diminui. Quanto maior o número de leis estabelecidas, e quanto mais fortes elas

[9] Esta afirmação solta precisa de muita elaboração. Espera-se que isto seja alcançado numa fase posterior. Temos em mente que as línguas L. referem-se as mesmas grandezas físicas sem, no entanto, existir uma traduzibilidade frase a frase entre elas.

são, menos eficiente se torna a linguagem. É plausível que com uma diminuição contínua da eficiência de uma língua, um estágio pode ser alcançado onde esta linguagem será completamente abandonada e substituída por outra que, na mesma evidência, mostra uma maior eficiência, principalmente através do fato de que as leis empíricas (ou pelo menos algumas) da primeira linguagem levaram a uma modificação do quadro conceitual da segunda.

> O nova iorquino, em nossa ilustração anterior, faria bem, depois de ter permanecido por algum tempo em San Francisco, ao adotar uma nova linguagem em que 'W' significaria mais de 60 °, "T" entre 50 ° e 60 ° e "C" para menos de 50, por exemplo, para salvá-lo de tornar ineficiente e desinteressante declarações sobre o tempo em San Francisco, que tinha antes em quase todos os casos, a forma 'T (x)', i. e., 'É temperado no tempo x'.

Às vezes pode ser útil falar sobre a ineficiência de uma linguagem. A definição é óbvia:

$$inef(L_1, in, e) =_{Df} 1 - ef(L_1, in, e).$$

É aconselhável evitar o termo 'redundância' - o termo usado para o correlato de nossa 'ineficiência' na teoria da comunicação - já que a expressão 'redundância de um quadro conceitual' é geralmente entendida em um sentido diferente.

§ 13. Conclusões

> Os conceitos de informação e medidas de informação explicadas aqui devem ser de valor em várias teorias, como na Teoria do Design de Experimentos e na Teoria do Teste. Várias extensões são descritas. Uma delas levaria em conta os arranjos lineares dos indivíduos.

A Teoria da Informação Semântica descrita aqui nada mais é do que uma certa ramificação da Teoria da Probabilidade Indutiva apresentada em [Prob.]. A explicação do conceito pré-sistemático da informação transportada por uma sentença, que foi tentada aqui, deve ser de valor para uma clarificação das fundações de todas aquelas teorias que fazem uso deste conceito e as medidas relacionadas a ele. O impacto dos conceitos aqui apresentados para a Teoria do Desenho de Experimentos ou para a teoria do teste deve ser óbvia[10].

[10] Veja, por exemplo, CRONBACH (1952).

A presente teoria requer extensão em várias direções. Uma extensão já foi mencionada: não há grandes dificuldades envolvidas no tratamento de sistemas linguísticos com denumeravelmente muitos indivíduos. Nem a introdução de variáveis individuais e a quantificação sobre elas apresentam problemas dos quais não sabemos a solução. Sistemas linguísticos desses tipos já foram tratados em [Prob.]

Outras extensões, no entanto, terão que ser adiadas até que as teorias correspondentes de probabilidade indutiva sejam desenvolvidas. Um de nós (R. C.) está empenhado, no momento, em desenvolver conceitos com a ajuda de quais sistemas linguísticos permitem levar em conta que possam ser investigados arranjos lineares de seus indivíduos. Isto sendo realizado, a construção de uma Teoria da Informação Semântica para universos sequenciais deve provar ser uma tarefa bastante simples.

Corrigenda

1. Substitua 'mínimo' por 'nulo' em T3-7, 9, 10 e 11.
2. Cancele T3-12 e o parágrafo seguinte.
3. Substitua o texto após D4-1 ao final da seção 4 pelo seguinte:

Vamos afirmar apenas um teorema sobre o conteúdo relativo:

T4-4.

a. Se i é uma sentença L-verdadeira, $Cont(j/i) = Cont(j)$.

Prova: Neste caso, $i \wedge j$ é equivalente a L para *j*. O teorema segue de D1 e T2*a*.

b. $Cont(j/t) = Cont(j)$

Assim, o conteúdo relativo de *j* em relação a *t* é igual ao conteúdo absoluto de *j*.

Portanto, seria possível começar com o conteúdo relativo como primitivo e definir o conteúdo absoluto como o valor do conteúdo relativo em relação a *t*. No entanto parece mais conveniente começar com o conceito simples de conteúdo absoluto, porque tem apenas um argumento e o conteúdo relativo pode ser definido com base nele. Este é o procedimento que escolhemos aqui.

Corrigenda adicional – Relatório Técnico n. 247

1. página 26, Tabela III, ao invés de $(n-1)/n$, leia-se $1-(1/2^n)$.

2. página 28, Eq. 2 Ao invés de $c_1(j)$, leia-se $c_1(i)$.

3. página 28, linha 7 de baixo para cima. Ao invés de T1b, leia-se T4b.

4. Página 31, Tabela V. Transponha os valores das colunas 2 e 3.

Tradução de Ralph Leal Heck[11]

Referências

CARNAP, R. 1942. *Introduction to Semantics*. Harvard University Press.

CARNAP, R. 1950. *Logical Foundations of Probability*. Chicago: University of Chicago Press.

CARNAP, R. 1952a. *The Continuum of Inductive Methods*. Chicago: University of Chicago Press.

CARNAP, R. 1952b. Meaning Postulates. P*hilosophical Studies*: An International Journal for Philosophy in the Analytic Tradition vol. 3 n.5, pp. 65 – 73. Doi: https://doi.org/10.1007/BF02350366

CRONBACH, L. J. 1952. *A Generalized Psychometric Theory Based on Information Measure*. A Preliminary Report. University of Illinois.

HOCKETT, C. F. An Approach to the Quantification of Semantic Noise. *Philosophy of Science*. vol. 19, n. 4, pp. 257-260.

SHANNON, C. E. 1948. Mathematical Theory of Communication. *Bell System Technical Journal* vol. 27, n. 3, pp. 379–423. Doi: https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

[11] Professor da Faculdade Católica de Fortaleza. Agradeço à professora Karen Leal Heck, que auxiliou na tradução e na revisão do texto.

Resumo

Neste artigo, Bar-Hillel e Carnap apresentam pela primeira vez as ideias básicas de uma teoria da informação semântica. O conceito de informação, baseado em sentença declarativa em uma dada linguagem, é definido pelo conteúdo descritivo expresso pelas sentenças. A quantidade de informação é calculada com base em um conjunto de funções lógicas probabilísticas sobre conjuntos de conteúdos descritíveis. Esta teoria tem aplicabilidade dedutiva e indutiva, significando um marco para o desenvolvimento das teorias filosóficas da informação.

Palavras-chave: teoria da informação semântica.

Abstract

In this article Bar-Hillel and Carnap present for the first time the basic ideas of a theory of semantic information. The concept of information, based on a declarative sentence in a given language, is defined by the descriptive content expressed by the sentences. The amount of information is calculated based on a set of probabilistic logic functions on sets of describable contents. This theory has both deductive and inductive applicability, being a milestone for the development of philosophical theories of information.

Keywords: *semantic information theory; logic; language; probability; logical space.*