# APÈNDIX 6: INFORMACIÓ

## LA TEORIA DE LA INFORMACIÓ

Suposem que tenim un sistema *S* amb un espai probabilístic d'esdeveniments $\Omega=\{S_1, S_2,...\}$, on els $S_i$ formen una partició de $\Omega$. Com podem definir la *informació* que hi ha en un esdeveniment $S_i$? Si aquest esdeveniment és altament probable, el seu coneixement no augmentarà significativament la nostra informació (l'anterior és el que ocorre, si sabem que en una escola de Finlàndia ningú parla la llengua catalana). Altrament, si l'esdeveniment $S_i$ és molt improbable, la informació adquirida serà altament significativa (això passa, si ens diuen que en un poblat de Groenlàndia trobem molts d'orangutans). Per tant, la informació que ens dóna el coneixement d'un esdeveniment tindrà una relació funcional decreixent amb la seva probabilitat. Donat que $0 \le P(S_i) \le 1$, la funció $-\log_a p(S_i)$ verifica la condició anterior i definirem la informació que assignem al *coneixement* d'un esdeveniment així:

$$INF(S_i) = -\log_a P(S_i)$$

La *informació mitjana H*, trobada a partir d'una *partició* $S_i$ de l'espai d'esdeveniments $\Omega$, s'anomena *entropia*:

$$H(S) = -\sum_i p(S_i).\log_a p(S_i)$$

Es demostra que $H \le \log_a n$, on *n* és el nombre d'esdeveniments possibles. Amb els esdeveniments equiprobables $H=\log_a n$ i tenim, tret d'un factor de proporcionalitat, l'entropia termodinàmica: *si creix n, augmenta la nostra ignorància sobre un sistema desconegut i la informació mitjana posterior al seu coneixement.*

Quan la base logarítmica és *2*, la informació es mesura en *bits.* La informació d'un esdeveniment amb probabilitat igual a *1/2* o la informació mitjana corresponent a un conjunt de dos esdeveniments equiprobables és d'*1* bit. En el que segueix suposarem que la base logarítmica és *2* i escriurem només *log*.

Amb una *font d'informació* que emet una seqüència de símbols, ens interessa conèixer les probabilitats en què cada símbol apareix. Una font d'informació de *memòria nul·la* és aquella en què els diferents símbols són estadísticament independents.

En una font d'informació de *Markov d'ordre m* la probabilitat d'emissió d'un símbol depèn dels *m* símbols que l'han precedit. Aquesta font ve determinada per les probabilitats condicionals

$$P(S_i / S_{j_1},...,S_{j_m})$$

Les fonts d'informació de *Markov* més importants són les *ergòdiques*. En una font ergòdica apareixen, a partir d'un temps suficientment llarg seqüències típiques de *m* símbols amb probabilitats concretes

$$P(S_{j_1},...,S_{j_m})$$

En una font d'informació markoviana ergòdica podem definir la *informació mitjana o entropia* d'una manera semblant al que hem fet abans:

$$H(S) = -\sum P(S_{j_1},...,S_{j_{m1}}).\sum P(S_i / S_{j_1},...,S_{j_{m1}}).\log P(S_i / S_{j_1},...,S_{j_m}) =$$

$$= -\sum P(S_{j_1},...,S_{j_{m1}},S_i).\log P(S_i / S_{j_1},...,S_{j_m})$$

L'estructura d'un determinat llenguatge, com el català, no es troba en l'emissió de símbols que fa una font sense memòria, ja que l'aparició de les diferents lletres de l'alfabet ve condicionada per les que l'han precedida. És per aquesta raó que a partir d'una font de *Markov* podem obtenir seqüències de lletres més representatives d'aquell llenguatge.

En l'emissió d'informació és essencial la creació d'una codificació no ambigua i el més curta possible. Cada símbol pot ésser codificat a partir de símbols binaris. El nombre de dígits binaris (*binits*) corresponents a un símbol defineix la seva longitud. Les probabilitats d'emissió de cada símbol fixaran la longitud mitjana dels símbols emesos per la font. Amb *n* binits podem codificar $2^n$ símbols. Si aquests símbols són equiprobables tindrem una infor-

mació mitjana de *n* bits. Tanmateix, si els símbols no són equiprobables, el nombre fix de binits emesos per una font per a cada símbol no coincidirà amb el nombre de bits.

Donada una distribució probabilística de l'emissió dels diferents símbols ens podem preguntar quina és la longitud mitjana en binits dels símbols emesos i també quina és la longitud mitjana mínima d'una codificació òptima. *El primer teorema de Shannon* ens contesta a la darrera pregunta tot afirmant que sota determinades condicions el nombre mitjà mínim de binits d'una font coincideix amb el valor de l'entropia *H(S).*

Suposem ara el sistema *S* (amb l'emissió de la variable aleatòria corresponent al conjunt $\{S_1,S_{2,...}\}$) en interacció amb un altre *A* (amb la recepció de la variable aleatòria $\{A_1,A_2,...\}$). La informació transmesa a través del *canal* ve determinada per les probabilitats $P(S_j/A_i)$. Aquestes probabilitats ens parlen de la fiabilitat de la transmissió i ens diuen quina és la probabilitat que *havent rebut el símbol* $A_i$ *s'hagi en realitat emès* $S_j$. Quan per a cada $S_j$ una d'aquestes probabilitats val *1* i la resta *0*, podem conèixer el missatge emès sense cap mena d'ambigüitat.

Podem definir les informacions mitjanes que segueixen:

$$H(S/A_i) = -\sum_j P(S_j/A_i).logP(S_j/A_i)$$

$$H(S/A) = -\sum_i \sum_j P(A_i).P(S_j/A_i).logP(S_j/A_i)$$

$$H(S,A) = -\sum_i \sum_j P(S_i \cap A_j).logP(S_i \cap A_j)$$

D'aquí arribem a

$$I(S,A) = H(S) - H(S/A) = I(A,S) = H(A) - H(A/S) \geq 0$$

i podem fer-ne les interpretacions que segueixen:

a) En el cas d'una comunicació completament fiable amb un canal sense sorolls el coneixement de la sortida ens dóna un coneixement total de l'entrada. En aquest cas $H(S/A)=0$ i $I(S,A)=H(S)$, mentre que quan perdem informació $H(S/A)>0$. $H(S)$ representa la informació de *S* i $H(S/A)$ la informació de *S* que no circula entre *S* i *A* (podem fer afirmacions semblants sobre $H(A)$ i

*H(A/S)*). Aquesta igualtat, complementària de les dels intercanvis de matèria i d'energia, representaria *l'intercanvi d'informació I(S,A)=I(A,S)* entre *S* i *A*.

b) Si *S* és un sistema biològic i *A* el seu ambient, *H(S) i H(A)* ens donarien una mesura de les seves complexitats pròpies i *H(S/A)* i *H(A/S)* la de les seves complexitats de resposta davant dels canvis de l'altre.

c) Si *S* i *A* són parts d'un sistema tenim aquests dos casos extrems:

1) *S* i *A* són totalment independents, això implica que *H(S/A)=H(S)* i *I(S,A)=0.* Aquesta situació correspondria a un sistema totalment caòtic i *s*ense comunicació entre les parts: tindríem *un desordre total*.

2) Totes les parts del sistema tenen un únic esdeveniment. En aquest cas també es verificaria *I(S,A)=0,* hi hauria *el màxim ordre* i totes les parts actuarien segons una única opció amb una rigidesa total del sistema.

La *informació algorísmica* que cal per descriure el caos és gran, mentre que la necessària per descriure l'ordre rígid és petita. La *complexitat* estructural és, però, molt diferent de la complexitat algorísmica i apareix entre els dos extrems anteriors, amb la qual cosa *I(S,A)* podria ser un bon índex per al coneixement de la complexitat d'un sistema.

La fiabilitat de la transmissió a través d'un canal variarà en funció de les probabilitats $P(S_i)$ de la font primària d'informació i la seva *capacitat* vindrà donada pel valor màxim de *I(S/A).*

*El segon teorema de Shannon* ens dóna les condicions perquè la probabilitat d'error en la interpretació de la sortida d'un canal sigui més petita que qualsevol nombre positiu $\varepsilon$ per petit que sigui. La *teoria dels codis correctors d'errors* apareix a partir de l'afirmació anterior. A hores d'ara encara no s'ha trobat una codificació ideal corresponent al resultat del teorema esmentat.

*La longitud i la fiabilitat* d'un missatge s'han d'optimitzar. El primer i el segon teoremes de *Shannon* fan afirmacions sobre el primer i el segon dels aspectes anteriors, respectivament. La in-

formació ha de tenir sovint, tanmateix, la propietat de la *confidencialitat*. La *criptografia* intenta preservar aquesta darrera. Les tècniques criptogràfiques més depurades permeten a tothom amb una *clau pública* encriptar i transmetre la informació, mentre que només qui posseeix la *clau privada* pot realitzar la desencriptació d'un missatge. El sistema *RSA* (acrònim dels seus inventors *Rivest, Shamir i Adelman*) respon a l'esquema anterior: un nombre natural molt gran facilita la seva clau pública, mentre que els dos nombres primers en què ell es descompon faciliten la seva clau privada. La dificultat d'aquesta descomposició factorial preserva la confidencialitat del missatge.

La *informació algorísmica* ve fixada pel grau *d'aleatorietat*: així, la seqüència *111111...* no conté informació algorísmica, mentre que una seqüència totalment aleatòria en conté molta. La *complexitat estructural,* però, és màxima per a un valor intermedi de la informació algorísmica, en què podem maximitzar el nombre de *regularitats* presents. L'univers primigeni *d'entropia física* baixa contenia poca informació algorísmica. Amb la seva evolució va anar augmentant l'entropia i la informació algorísmica i es formaren estructures complexes. Malgrat que amb el temps l'entropia de l'univers serà massa gran, localment podem tenir un contingut entròpic i informatiu limitat, perquè les estructures complexes es puguin continuar formant. Mentre *no estiguem molt a prop de l'equilibri* final de l'univers, serà possible (d'acord amb *Prigogine*) obtenir *localment* una complexitat creadora a partir de *bifurcacions* degudes a *fluctuacions* amplificades gràcies a la *no-linealitat* gravitatòria. Les *estructures locals dissipatives* es mantindrien, en part, gràcies a la seva interacció *informativa* amb la resta del món i la creació de la complexitat i l'evolució podrien esdevenir processos on *el tractament de la informació* fos essencial.

Aquí ens cal puntualitzar que la *informació algorísmica o complexitat KC (Kolmogorov-Chaitin)* és diferent de *la complexitat computacional* que ve definida pel temps necessari per solucionar un problema. Així, la descomposició d'un nombre enter en producte de dos de primers no té gairebé complexitat algorísmica, però sí complexitat computacional; en efecte: el temps necessari per fer-ne la descomposició no creix *potencialment* amb la longitud

del nombre (problema *tractable*), sinó *exponencialment* (problema *intractable*).

El temps necessari per desenvolupar una computació entorn de la descripció d'un sistema origina la seva *complexitat computacional*. La complexitat computacional de la totalitat en funció de les parts defineix la seva *profunditat* i la corresponent a la descripció de les parts en funció de la totalitat la seva *cripticitat*. Així, la descomposició d'un nombre molt gran en producte de dos de primers té molta cripticitat i poca profunditat, mentre que l'estructura bioquímica de la vida té molta profunditat (és molt complex explicar la vida a partir dels seus components) i menys cripticitat (és relativament més senzill trobar l'estructura atòmica de la vida).

Segons el principi de *Landauer* és necessària una energia mínima per esborrar part de la informació d'un ordinador, energia que portaria aparellat un augment entròpic. Això vindria a ser l'equivalent a les limitacions pròpies del rendiment de les màquines tèrmiques i del tractament de la informació, que varen teoritzar *Carnot* i *Shannon*, respectivament. Tanmateix, *Bennet* ha demostrat que aquell esborrat no és obligatori i evita la despesa energètica mitjançant la reversibilitat de les operacions de la computabilitat; només si es vol repetir el procés cal fer-ne l'esborrat amb la despesa energètica i el augment entròpic corresponents. És aquest esborrat cíclic el que permet resoldre la paradoxa del *dimoni de Maxwell* que es tracta a l'apèndix 5 de la termodinàmica.

*Wheeler* amb la seva frase "*it from bit"* afirmava la influència de la informació en les lleis físiques del món i l'univers no seria altra cosa que un procés d'informació, una fabulosa computació del Gran Simulador de *Deutsch.* El creixement entròpic de l'univers portaria aparellat un augment de la seva informació que faria possible "el factor sorpresa" amb l'aparició de la novetat al món. L'anterior potser seria possible a través de successives *iteracions* computacionals que donarien lloc a l'aparició de nous atractors fractals amb les seves corresponents estructures evolutives de complexitat creixent.

Si la influència física de la informació fos certa, la pregunta de *Wheeler* "*Why quantum*?" tindria una resposta possible i provisional: el nostre coneixement de la realitat física seria quàntic, perquè ho seria la informació a través de la qual ella se'ns manifesta i que és discreta en estar formada per bits.

# ELS PROCESSOS DE MARKOV

Un procés general (no necessàriament de *Markov*) governa les fluctuacions entre diferents estats a través d'una equació evolutiva que relaciona la derivada temporal de la probabilitat de cada estat amb les probabilitats de transició des d'aquest estat i cap a aquest estat. D'acord amb això, la probabilitat $P(S_i)$ d'un estat en un instant varia amb el temps. Si tenim un procés de *Markov* d'ordre *n,* la probabilitat que aparegui en l'evolució la seqüència d'estats $S_aS_bS_c...S_hS_i$ *serà* $P(S_aS_bS_c...S_h).P(S_i|S_c...S_h)$, on la seqüència $S_c...S_h$ conté únicament els *n* estats immediatament anteriors a $S_i$ i es perd el "record" de tots els altres.

En *un procés de Markov de primer ordre* es verificarà

$$\frac{dP(S_i)}{dt} = \sum_{S_i \neq S_j} (P(S_j,t).P(S_j \to S_i) - P(S_i,t).P(S_i \to S_j)$$

, on $P(S_i \to S_j)$ és la probabilitat de transició de $S_i$ a $S_j$ per unitat de temps. L'equació anterior s'anomena *equació Master* del procés de *Markov* i *no és invariant per inversió temporal*.

En *un procés de Markov ergòdic* cada estat és visitat amb probabilitat *1* a través d'un temps mitjà finit. Com a conseqüència, es verifica que $\Delta H(S) \geq 0$ i existeix una distribució estacionària amb $dP(S_i)/dt=0$, que satisfà l'equació Master i a la qual tendeixen totes les solucions d'aquesta amb el temps. Un procés de *Markov* marca, doncs, una clara *asimetria temporal* fins arribar a la solució estacionària. Pot, però, un procés de *Markov* sorgir tot admetent les equacions reversibles de la dinàmica? La resposta, sota determinades hipòtesis *restrictives*, és afirmativa. La reversibilitat, a través d'una transformació en què s'explicita la no-localitat, pròpia de la coherència espacial, i es prescindeix de la informació no manifesta, tot integrant-la, desemboca en un procés

de *Markov* amb el pas de les col·lectivitats puntuals *de Gibbs*, de precisió infinita, a les no locals de *Boltzmann,* de precisió finita. No és, però, il·lícit fer aquest pas i no estem renunciant a una precisió *teòrica* infinita de les condicions inicials *reals*? Ben al contrari. Un sistema dinàmic conservatiu seria altament inestable. L'única estabilitat possible seria l'estabilitat de *Liapunov* per a sistemes simples a les rodalies dels punts *el·líptics* (vegeu el capítol 3). Tanmateix, la presència *densa* en sistemes complexos de punts *hiperbòlics* dóna lloc a grans inestabilitats i l'indeterminisme quàntic intrínsec faria que finalment, *fins i tot dins d'un formulisme clàssic,* fos inevitable el pas des de les col·lectivitats de *Gibbs* cap a les col·lectivitats de *Boltzmann* amb la integració ans esmentada de part de la informació. De fet aquesta integració està implícita en l'aparició de les probabilitats a l'equació Master, tot ignorant les amplituds de probabilitat dels estats quàntics corresponents.

Quan estudiem un sistema amb *N* estats a través d'un procés de *Markov* de *m* passos obtenim $N^m=exp(m.lnN)$ seqüències possibles de longitud *m* que defineixen una *cadena de Markov.* Si ordenem les seqüències per probabilitats decreixents, el número de seqüències amb la suma de probabilitats propera a *1* és

$$n_m \cong exp(m.H)$$

, on

$$H = -\sum_{ij} P(S_i).P(S_j|S_i).ln\ P(S_j|S_i)$$

és *l'entropia associada a la cadena de Markov.*

El valor màxim de *H* ocorre en l'equilibri, on tots els estats són equiprobables, i val $H_{max}=lnN.$ En aquesta situació totes les seqüències tindrien la mateixa probabilitat d'aparèixer i, de fet, cada seqüència concreta seria altament improbable.

Per a situacions molt allunyades de l'equilibri $H<<lnN$ i un procés de *Markov* actuarà *seleccionant* una classe molt concreta de seqüències. Aquestes seqüències tindrien, doncs, una probabilitat important d'aparèixer i serien significatives: és el que ocorre en els processos cel·lulars de síntesi, on només es manifesten aquelles proteïnes, formades a partir dels vint tipus d'aminoàcids, que tenen funcions específiques i essencials.