A Framework for representing knowldege*

Marvin Minsky

MIT-AI Laboratory Memo 306, June, 1974.

1. Marcos

Me parece que los ingredientes de la mayoría de las teorías tanto en la Inteligencia Artificial como en la Psicología han sido en general demasiado menudas, locales, y no estructuradas para explicar — tanto practica o fenomenológicamente — la efectividad del pensamiento con *sentido-común*. Los "pedazos" de razonamiento, lenguaje, memoria, y "percepción" deben ser más grandes y estructurados; su contenido basado en hechos y procedimientos debe conectarse más íntimamente para poder explicar el indudable poder y velocidad de las actividades mentales.

Sensaciones similares parecen estar emergiendo en varios centros que trabajan en teorías de la inteligencia. Toman forma en la propuesta de Pappert y yo (1972) para sub-estructurar el conocimiento en micro-mundos; otra forma en los "Espacios de problemas" de Newell y Simon (1972); y otra más en nuevas estructuras mayores que teóricos como Schank (1974), Abelson (1974), y Norman (1972) asignan a los objetos lingüísticos. Veo a todos éstos apartándose de los intentos tradicionales de los psicólogos conductistas y de los estudiantes de Inteligencia Artificial orientados a la lógica, tratando de representar conocimiento como una colecciones de fragmentos separados, sencillos.

Aquí trato de juntar varias de estas ideas pretendiendo obtener una teoría unificada y coherente. El escrito plantea más preguntas que las que responde, y he tratado de observar las deficiencias de la teoría.

Aquí esta la esencia de la teoría: Cuando uno encuentra una nueva situación (o hace un cambio substancial en la vista de uno del problema presente) uno selecciona de la memoria una estructura llamada *marco*. Esta es un marco de referencia recordado para ser adaptado para ajustarse a la realidad cambiando los detalles conforme sea necesario.

Un *marco* es una estructura de datos para representar una situación estereotipada, como estar en cierta clase de salón, o ira a una fiesta de cumpleaños de un niño. A cada marco están asociados varios tipos de información. Alguna

*Reprinted in The Psychology of Computer Vision, P. Winston (Ed.), McGraw-Hill, 1975. Shorter versions in J. Haugeland, Ed., Mind Design, MIT Press, 1981, and in Cognitive Science, Collins, Allan and Edward E. Smith (eds.) Morgan-Kaufmann, 1992 ISBN 55860-013-2]

de esta información es acerca de como usar el marco. Alguna es acerca de qué puede uno esperar que ocurra después. Alguna es acerque de qué hacer si esas expectativas no se confirman.

Podemos pensar de un marco como una red de nodos y relaciones. Los "niveles superiores" de un marco están fijos, y representan cosas que siempre son verdaderas acerca de una situación supuesta. Los niveles inferiores tienen muchas *terminales* —"ranuras" que deben llenarse con casos específicos o datos. Cada terminal puede especificar condiciones que deben cumplir sus asignaciones. (Las asignaciones mismas usualmente son "sub-marcos" más pequeños.) Las condiciones simples se especifican por *marcadores* que pueden requerir que una asignación terminal sea una persona, un objeto de suficiente valor, o una apuntador a un sub-marco de cierto tipo. Condiciones más complejas pueden especificar relaciones entre las cosas asignadas a varias terminales.

Las colecciones de marcos relacionados se encadenan en *sistemas de marcos.* Los efectos de las acciones importantes son reflejadas por *transformaciones* entre los marcos de un sistema. Estas se usan para hacer más económicos ciertos tipos de cálculos, para representar cambios de énfasis y atención, y para explicar la efectividad de "imaginería."

Para el análisis de la escena visual, los diferentes marcos de un sistema describen la escena desde distintos puntos de vista, y las transformaciones entre un marco y otro representan los efectos de moverse de lugar a lugar. Para las clases de marcos no-visuales, las diferencias entre los marcos de un sistema pueden representar acciones, relaciones causa-efecto, o cambios en el punto de vista conceptual. *Diferentes marcos de un sistema comparten las mismas terminales*; este es el punto critico que hace posible el coordinar información acumulada desde distintos puntos de vista.

El poder fenomenológico se debe en mayor parte a que la teoría se articula en la inclusión de expectativas y otros tipos de presunciones. *Las terminales de un marco normalmente se ya están llenas con asignaciones preestablecidas.* De aquí que un marco pueda tener gran cantidad de detalles cuya suposición no está garantizada específicamente por la situación. Estas tienen muchos usos en la representación de información general, casos más posibles, técnicas para evitar la "lógica," y maneras de hacer generalizaciones útiles.

Las asignaciones preestablecidas están conectadas débilmente a sus terminales, de tal forma que pueden ser desplazadas por nuevos objetos que mejor se ajusten a la situación actual. Entonces también pueden servir como "variables" o como casos especiales de "razonamiento por ejemplo," o como "casos de libro de texto," y frecuentemente hacen innecesario el uso de cuantificadores lógicos.

Los sistemas de marcos son ligados, a su vez, por una *red de recuperación de información.* Cuando un marco propuesto no puede ajustarse a la realidad — cuando no podemos encontrar asignaciones terminales que concuerden adecuadamente con las condiciones de su marcador terminal — esta red provee un marco de reemplazo. Esas estructuras entre-marcos posibilitan otras formas de representar conocimiento acerca de hechos, analogías, y otra información útil en el entendimiento.

Una vez que un marco es propuesto para representar una situación, un proce-

so de *concordancia* trata de asignar valores a cada terminal del marco, consistente con los marcadores en cada lugar. El proceso de concordancia esta controlado en parte por la información asociada con el marco (la cual incluye información acerca de como tratar con sorpresas) y en parte por el conocimiento acerca de los objetivos actuales del sistema. Cuando un proceso de concordancia falla, hay un uso importante de la información obtenida. Discutiré como puede ser usada para seleccionar un marco alternativo que mejor se ajuste a la situación.

¡Disculpa! Los esquemas aquí propuestos están incompletos en muchos sentidos. Primero, frecuentemente propongo representaciones sin especificar los procesos que los usaran. Algunas veces solo describo las propiedades que deben exhibir las estructuras. Hablo de los marcadores y asignaciones como si fuese obvio como son asociados y ligados; no es así.

Aparte de las lagunas técnicas, hablaré como si no estuviera advertido de muchos problemas relacionados con el "entendimiento" que realmente necesita un análisis más profundo. No clamo que las ideas propuestas aquí sean suficientes para una teoría completa, solo que el esquema de los sistemas de marcos puede ayudar a explicar varios fenómenos de la inteligencia humana. La idea básica de marco en si no es particularmente original — esta en la tradición de "esquema" de Bartlett y de "paradigmas" de Khun; la idea de sistema de marcos es probablemente más novedosa. Winograd (1974) discute la tendencia actual, en teorías de la Inteligencia Artificial, hacia las ideas afines a marcos.

El resto de 1 aplica la idea de los sistemas de marcos a la visión e imágenes. En 2 cambiamos a la lingüística y otros tipos de entendimiento. 3 discute la memoria, adquisición y recuperación de conocimiento; 4 es acerca de control, y 5 toma otros problemas de visión e imágenes espaciales.

En el cuerpo de este escrito discuto una variedad de clases de razonamiento por analogía, y formas de imponer estereotipas a la realidad y saltar a conclusiones basadas en la concordancia parcialmente similar. Estos métodos son básicamente inciertos. ¿ Por qué no usar métodos más "lógicos" y certeros? La sección 6 es una suerte de apéndice que argumenta que la lógica tradicional no puede tratar muy bien con los problemas realistas, y complicados porque esta pobremente adecuada para representar *aproximaciones* a soluciones — y éstas son absolutamente vitales.

> *El pensar siempre comienza con planes e imágenes sugestivos pero imperfectos; éstos se reemplazan progresivamente por mejores ideas — aunque por lo común todavía imperfectas.*

6. Apéndice: Critica al enfoque logístico

> *"Si uno trata de describir procesos de pensamiento genuino en términos de la lógica formal tradicional, el resultado es con frecuencia insatisfactorio, entonces, uno tiene una serie de operaciones correctas, pero el sentido de el proceso que fue vital, contundente, creativo en este parece de alguna forma haberse evaporado en las formulaciones."*
>
> — *N. Wertheimer [Productive Thinking]*

Aquí explico porque pienso que los enfoques más "lógicos" no funcionarán. Han habido serios intentos, tan antiguos como Aristóteles, para representar el razonamiento por sentido común por un sistema "logístico" — esto es, uno que hace una separación completa entre

(1) "proposiciones" que engloban información específica, y

(2) "silogismos" o leyes generales para una inferencia apropiada.

Nadie ha sido capaz de afrontar exitosamente tal sistema con un conjunto realmente grande de proposiciones. Pienso que tales intentos seguirán fallando por el carácter de logístico en general, y no por los defectos de los formalismos particulares. (Intentos más recientes han usado variantes de la "lógica de predicados de primer orden," pero no pienso que *ese* sea el problema.)

Un intento típico por los sistemas logísticos para simular el razonamiento de sentido común comienza en un "micro-mundo" de complicación limitada. Por un lado se tienen como fin objetivos de alto nivel tales como "Yo quiero ir desde mi casa al aeropuerto." Por otro lado comenzamos con muchos objetos pequeños — los *axiomas* — como "el carro esta en la cochera," "uno no sale sin vestir," "para ir de un lado, uno debe moverse (completo) en su dirección," etc. Para hacer que funcione el sistema uno diseña procedimientos heurísticos de búsqueda para "demostrar" el objetivo deseado, o para producir una lista de acciones que lograran el resultado.

No voy a hacer un recuento de la historia de intentos para hacer coincidir ambos fines — pero solamente sintetizaré mi impresión: en casos simples uno puede hacer que tales sistemas "funcionen," pero conforme nos aproximamos a la realidad los obstáculos se vuelven abrumadores. El problema de encontrar los axiomas adecuados — el problema de "establecer los hechos" en términos de asunciones lógicas, siempre-correctas, es más difícil de lo que generalmente se cree.

Formalizar el conocimiento requerido: Solamente construir una base de conocimiento es un problema intelectual importante de investigación. Ya sea que el objetivo de uno sea logístico o no, todavía sabemos muy poco acerca del contenido y estructura del conocimiento del sentido común. Un sistema "mínimo" de sentido común debe "saber" algo acerca de causa y efecto, tiempo, propósito, localidad, proceso, y tipos de conocimiento. También necesita maneras de

adquirir, representar, y usar tal conocimiento. Necesitamos un serio esfuerzo de investigación epistemológica en esta área. Los ensayos de McCarthy[] y [] Sandewall son pasos en esa dirección. No tengo ningún plan fácil para esta enorme empresa; pero la magnitud de la tarea ciertamente dependerá fuertemente de las representaciones seleccionadas, y pienso que esa logística ya esta causando problemas.

Relevancia: ¡El problema de seleccionar relevancia de una variedad excesiva es un resultado clave! ¡Una epistemología moderna no se parecerá a las antiguas! Los conceptos computacionales son necesarios y novedosos. Tal vez la mejor parte del conocimiento no es "proposicional" en carácter, más bien interproposicional. Para cada "hecho" uno necesita meta-hechos acerca de como será usado, y cuando no debe ser usado. En el paradigma del "Aeropuerto" de McCarthy vemos formas de tratar con algunas interacciones entre "situaciones, acciones, y leyes causales" dentro de un micro-mundo restringido de cosas y acciones. Pero mientras el sistema puede hacer deducciones implicadas por sus axiomas, pero no puede decírsele cuando *debe* o no debe hacer tales deducciones.

Por ejemplo, uno puede querer decirle al sistema que "no cruce la calle si viene un coche." Pero uno no puede pedirle al sistema que "demuestre" que no viene ningún coche, normalmente no habrá tal demostración. En `PLANNER`, uno puede dirigir un *intento* para demostrar que viene un coche, y si el intento de deducción (limitada) finaliza con "falla," uno puede actuar. Esto no puede hacerse un sistema logístico puro. "Mira a la derecha, mira a la izquierda" es una primera aproximación. Pero si uno le dice al sistema la verdad real de las velocidades, los caminos cerrados, y las probabilidades de que autos de carrera pasen por la esquina, la demostración se vuelve poco práctica. Si lee en un libro de física que los campos intensos perturban los rayos de luz, ¿Deberá temer que un científico loco haya inventado un auto invisible? ¡Necesitamos representar lo "usual"! Eventualmente deberá entender la ventaja de hacerlo (cruzar la calle) contra el costo de la mortalidad, uno no puede paralizarse por el miedo.

Mortalidad: Aún si formulamos restricciones relevantes, los sistemas logísticos tienen problemas usándolas. En cualquier sistema logísticas, todos los axiomas son necesariamente "permisivos' — todos ayudan a que puedan derivarse nuevas inferencias. Cada axioma agregado significa más teoremas, ninguno puede desaparecer. ¡Simplemente no hay forma directa de agregar información para decirle a tal sistema acerca de los tipos de conclusiones que *no* deben ser deducidas! Para plantearlo sencillamente: si adoptamos suficientes axiomas para deducir lo que necesitamos, deducimos muchas otras cosas. Pero si tratamos de cambiar esto agregando axiomas acerca de la relevancia, todavía producimos todos los teoremas no deseados, e incomodas afirmaciones acerca de su irrelevancia.

Porque a los lógicos no les interesan sistemas que serán agrandados, pueden diseñar axiomas que permitan solo las conclusiones que quieren. En el desarrollo de la Inteligencia la situación es distinta. Uno tiene que aprender cuales caracte-

rísticas de las situaciones son importantes, y qué clase de deducciones no deben tomarse seriamente. La reacción usual a la "paradoja de los mentirosos" es, después de un rato, reír. ¡La conclusión es no rechazar un axioma, sino rechazar la deducción! Esto da lugar a otra cuestión:

Procedimiento para controlar el conocimiento: La separación entre axiomas y deducción hace poco práctico el incluir conocimiento que clasifique el conocimiento acerca de las proposiciones. Tampoco podemos incluir conocimiento acerca del manejo de la deducción. Un paradigma del problema es el de axiomatizar los conceptos cotidianos de aproximación o cercanía. Uno querría que la cercanía fuese transitiva:

$$(A \texttt{ cerca-de } B) \,\mathrm{Y}\, (B \texttt{ cerca-de } C) \Rightarrow (A \texttt{ cerca-de } C)$$

pero una aplicación no restringida de la regla haría todo cercano a todo lo demás. Uno puede intentar trucos técnicos como:

$$(A \texttt{ cerca-de}^*1 \; B) \,\mathrm{Y}\, (B \texttt{ cerca-de}^*3 \; C) \Rightarrow (A \texttt{ cerca-de}^*1 \; C)$$

y admitir solamente (digamos) cinco grados de `cerca-de`^{*1}, `cerca-de`^{*2}, `cerca-de`^{*3}, etc. Uno puede inventar cantidades o parámetros análogos. Pero uno no puede (en un sistema logístico) decidir el hacer una nueva clase de "axioma" para prevenir la aplicación de la transitividad después de (digamos) tres usos encadenados, condicionada, a menos de que haya una "buena excusa." No quiero decir el proponer una solución particular a la transitividad de cercanía. (Hasta donde se, todavía nadie ha hecho una propuesta acreditada acerca de ella.) Mi queja es que por la aceptación del la logística, nadie ha explorado libremente esta clase de restricciones procedurales.

Problemas combinatorios: No veo razón para esperar que éstos sistemas escapen a la explosión combinatoria cuando les sea dada una base de conocimientos más rica. Aunque vemos que se alientan las demostraciones en micro-mundos, de tiempo en tiempo, es común en la investigación en Inteligencia Artificial encontrar un alto grado de rendimiento en rompecabezas difíciles — dada solo la información justa para resolver el problema — pero esto no siempre lleva a un buen rendimiento en dominios más grandes.

Consistencia y completes: Un pensador humano revisa planes y listas de objetivos mientras trabaja, revisando su conocimiento y las políticas acerca de su uso. Uno puede programar algunas de éstas en el programa demostrador de teoremas — pero realmente uno también quiere representarlo directamente, de manera natural, en el cuerpo declarativo — para usarse en introspección adicional. ¿Entonces por qué tendrían los trabajadores que intentar que los sistemas logísticos hagan el trabajo? Una razón válida es que los sistemas tienen una simple elegancia atractiva; si trabajaran ésto estaría bien. Una razón inválida se ofrece con más frecuencia: que tales sistemas tienen una virtud matemática porque son:

(1) Completos — "Todas las afirmaciones verdaderas pueden demostrarse"; y

(2) Consistentes — "Ninguna afirmación falsa puede demostrarse."

Parece que no siempre se tiene en cuenta que la completes no es un premio raro. Es una consecuencia trivial de cualquier procedimiento de búsqueda exhaustiva, y cualquier sistema puede "completarse" adicionando le cualquier sistema completo y entrelazando los pasos computacionales. la consistencia es más refinada; requiere que los axiomas no impliquen contradicciones. Pero yo no creo que la consistencia sea necesaria o aún deseable en el desarrollo de sistemas inteligentes. Nadie es siempre completamente consistente. Lo que importa es como uno maneja las paradojas o conflictos, como uno aprende de los errores, como uno se aparta de las inconsistencias sospechadas.

Porque esta clase de idea falsa, el Teorema de Incompletez de Gödel ha estimulado mucha insensatez respecto a las supuestas diferencias entre las máquinas y el hombre. Nadie parece haberse dado cuenta de su interpretación más "lógica": forzar la consistencia produce limitaciones. Por supuesto que habrán diferencias entre humanos (quienes son demostrablemente inconsistentes) y máquinas cuyos diseñadores les han impuesto consistencia. Pero no es inherente a las máquinas el que sean programadas solo con sistemas lógicos consistentes. ¡Éstas discusiones "filosóficas" hacen totalmente innecesaria esta asunción! (Respeto la demostración reciente acerca de la consistencia de la teoría moderna de conjuntos, de esta manera, como indicio de que la teoría de conjuntos probablemente no es adecuada para nuestros propósitos — ¡no como evidencia tranquilizadora de que es seguro usar la teoría de conjuntos!)

Un matemático famoso, advirtió que su demostración podría llevar a una paradoja si tomara un paso lógico adicional, respondió "Ah, pero no tomaré ese paso." Lo decía totalmente serio. Una gran parte del conocimiento ordinario (o aún el matemático) se asemeja al de las profesiones peligrosas: cuando algunas acciones son poco aconsejables. ¿Cuando son ciertas aproximaciones seguras de usar? ¿Cuando ciertas medidas llevan estimaciones sensatas? ¿Cuales afirmaciones auto-referenciadas son permitidas si no son llevadas muy lejos? Conceptos como "cercanía" son muy valiosos como para abandonarlos solo porque nadie puede exhibir axiomas satisfactorios para éstos. Para resumir:

1. El razonamiento "lógico" no es lo suficientemente flexible para servir como una base para el pensamiento; prefiero pensar de este como una colección de métodos heurísticos, efectivos solo cuando se aplican a planes esquemáticos cabalmente simplificados. La Consistencia que demanda terminantemente la Lógica no es usualmente disponible de otra manera — *¡ Y probablemente no sea ni siquiera deseable!* — porque los sistemas consistentes tienden a ser muy "débiles."

2. Dudo de la factibilidad de representar al conocimiento ordinario en la forma de muchos proposiciones pequeñas, independientemente "verdaderas."

3. La estrategia de separar completamente el conocimiento especifico de las reglas generales de inferencia es muy radical. Necesitamos maneras más direc-

tas para ligar los fragmentos de conocimiento que recomienden *como* deben ser usadas.

4. Por mucho tiempo se ha creído que era crucial el hacer que todo accesible todo el conocimiento a la deducción en forma de declaraciones afirmativas; pero esto parece menos urgente conforme aprendemos formas de manipular descripciones estructurales y procedurales.

No quiero sugerir que el "pensamiento" puede ir muy lejos sin algo como el "razonamiento." Con certeza necesitamos (y usamos) algo como la deducción silogística; pero espero mecanismos para hacer tales cosas para surgir en cualquier caso de procesos para la "concordación" e "instanciación" requeridas para las otras funciones. La lógica formal tradicional es una herramienta técnica para discutir ya sea *todo lo que pueda ser deducido de algunos datos* o, *cuando cierta consecuencia puede ser deducida así*; no puede discutirse en absoluto que *debe* ser deducido bajo circunstancias ordinarias. Como la teoría abstracta de la sintaxis, la lógica formal sin una semántica procedural poderosa no puede tratar con situaciones significativas.

No puedo afirmar lo suficientemente fuerte mi convicción de que la preocupación con la Consistencia, tan valiosa para la Lógica Matemática, haya sido increíblemente destructiva para quienes trabajan en modelos de la mente. Al nivel popular ha producido una rara concepción de las capacidades potenciales de las máquinas en general. En el nivel "lógico" ha bloqueado los esfuerzos para representar el conocimiento ordinario, presentando una inalcanzable imagen de un conjunto de "verdades" libres de contexto que puedan sostenerse casi por si mismas. Y al nivel del modelado intelectual ha bloqueado la comprensión fundamental de que el *pensamiento comienza con planes e imágenes sugestivos pero defectuosos, que son lentamente (si alguna vez) refinados y reemplazados por unos mejores.*