

語句の意味記述の前に文の意味記述を

(計算機用の)辞書に書かれている語句の意味と、
文を通じて実際に理解される語句の意味の間隙を埋めるために

黒田 航

情報通信研究機構 けいはんな情報通信融合研究センター

京大/NTT 合同研究会
2005/12/18 (Thu)

はじめに「お断り」

- 私はこの発表でそれなりに「過激」な主張をします。例えば、次のようなことを(暗に)主張します:
 - コトバの意味が何であるかを最初に定義/規定し、それによって記述(可能)対象を制限するのは本末転倒だ。なぜなら誰も意味が何であるのかを知らないのだから。
 - 特に特定の文法フォーマリズム (HPSG など) を尊重しすぎないことが必要だ
- これらの点に関して、皆さんから「叩かれる」のを覚悟で発表に来ました。どうかよろしく (^_^)

本日の発表の流れ

- これまでの成果物の簡単な紹介
 - 複層意味フレーム分析 (Multilayered Semantic Frame Analysis: MSFA) を用いた文の意味の直接記述/解析
 - MSFA の詳細は黒田・井佐原 (2004) を
- MSFA の開発方針, というか 設計“思想” の説明

簡単な成果紹介

私の研究の方向

- 私は次のことをしている, 事実上「元」言語学者
 - 意味役割タグづけによる言語資源の開発
 - 意味役割タグ体系の設計
 - タグづけ仕様の定義
 - 意味役割(名) (semantic role (name)s) と意味型(名) (semantic type (name)s) の区別に基づくシソーラスの構造の再検討 (と再構築?)
 - 意味役割に基づく概念化の一般理論を FOCAL (黒田・中本・野澤 2005) の名で構築
 - 従来の非科学的言語学からの脱却の促進

私の研究の方向

- 私は次のことをしている, 事実上「元」言語学者
 - 意味役割タグづけによる言語資源の開発
 - 意味役割タグ体系の設計
 - タグづけ仕様の定義
 - 意味役割(名) (semantic role (name)s) と意味型(名) (semantic type (name)s) の区別に基づくシソーラスの構造の再検討 (と再構築?)
 - 意味役割に基づく概念化の一般理論を FOCAL (黒田・中本・野澤 2005) の名で構築
 - 従来の非科学的言語学からの脱却の促進

意味役割タグづけの一例

- 京大コーパスから
 - 超大型FW同士の激闘が大学日本一決戦を彩ることになった。[S-ID: 950103083-001]
- この文にはメタファー (“彩る”) あり，複合表現 (“大学日本一決戦”) ありで，意味の記述は決して簡単ではない
 - <http://www.kotonoba.net/~mutiyama/cgi-bin/hiki2/hiki.cgi?c=view&p=msfa-round1-s01>

MSFA の実例集

- 京大コーパスから適当に選んだ文に MSFA で意味役割をタグづけした結果が閲覧制限つきで公開中
 - <http://www.kotonoba.net/~mutiyama/cgi-bin/hiki2/hiki.cgi?c=view&p=msfa-round3-s1>
 - <http://www.kotonoba.net/~mutiyama/cgi-bin/hiki2/hiki.cgi?c=view&p=msfa-round3-s2>
 - <http://www.kotonoba.net/~mutiyama/cgi-bin/hiki2/hiki.cgi?c=view&p=msfa-round3-s3>
 - <http://www.kotonoba.net/~mutiyama/cgi-bin/hiki2/hiki.cgi?c=view&p=msfa-round3-s4>
 - <http://www.kotonoba.net/~mutiyama/cgi-bin/hiki2/hiki.cgi?c=view&p=msfa-round3-s5>
 - <http://www.kotonoba.net/~mutiyama/cgi-bin/hiki2/hiki.cgi?c=view&p=msfa-round3-s6>
 - <http://www.kotonoba.net/~mutiyama/cgi-bin/hiki2/hiki.cgi?c=view&p=msfa-round3-s7>
 - <http://www.kotonoba.net/~mutiyama/cgi-bin/hiki2/hiki.cgi?c=view&p=msfa-round3-s8>
 - <http://www.kotonoba.net/~mutiyama/cgi-bin/hiki2/hiki.cgi?c=view&p=msfa-round3-s9>
- アクセスには私が発行するアカウントが必要

何のための意味記述か？

言語研究の基本に戻る

出発点

- ヒトの理解は状況依存的であり，意味型のレベルではなく意味役割レベルで達成される
 - 状況意味論 (Barwise and Perry 1983) で言う “状況” とは完全に同一ではない (ようだ)
 - 詳細は中本・黒田・野澤 (2005), 黒田ほか (2005a, b) を
- これが正しいとすれば，意味型分類を基本原則として構築された従来の辞書，シソーラス (e.g., 日本語語彙大系) や，それらを元にした意味記述の有用性には限界がある

基本方針 [1/3]

- 任意の文 s に有用な意味記述 $SD(s)$ を与える際
 - 文脈から独立した語の意味を仮定しない
 - 語句の意味の架空の実体性の回避
 - 文 $s = w_1 \cdot w_2 \cdots w_n$ を形式的に構成している語句の意味がすでに何からの形で「辞書」に書かれていると仮定しない
 - 特定の統語理論の妥当性を前提としない
 - 意味記述のための理論の統語理論からの独立性の確保
- $SD(s)$ は s の意味アノテーションに相当する
 - 意味記述は (少なくとも当面) 表示である必要はない

基本方針 [2/3]

- 私が強く主張すること
 - 意味記述の有用性は既成の統語理論 (e.g. HPSG, LFG) との整合性によって測られるべきではない
 - 「何が文の意味であるか」がそもそもわかっていない以上、文の意味表現に統語理論との整合性を要求するのは単なる「ない物ねだり」かも知れない
 - 一部の意味が統語構造と対応するからといって、それを理由に「言語の意味が何らかの形で、すべて統語構造に対応づけられる」と考えるのは言語学者の誇大妄想

基本方針 [3/3]

- 必要なのは語句の意味の“生態記述”
 - 理論的前提とアプローチの仕方は異なり，語句の意味を動的に“生成”する装置を開発するのではなく，その“生態”を直接記述する方法を開発するのが私の狙い
 - 語句の意味が本質的に流動的であることを理解し，その性質性を失わない記述を行なう必要があることを強調する点で，生成辞書理論 (Pustejovsky 1991, 1995) と問題意識を共有

コトバの“意味”とは何か？

- 二つの別の問いを混同しない:
 - A. 一般に，コトバの意味とは何か？
 - B. 特定の目標のために，コトバの意味をどう記述するか？
- 問い B の答えは目的に依存するが，問い A の答えはそうではない
 - 例えば，機械翻訳という課題で問題にすべきコトバの意味の範囲は，ヒトの理解可能性の全体の規定という課題で問題にするコトバの意味の範囲の一部

コトバの意味の“定義”??

- コトバの意味の「正しい定義」を求めているのは誰?
- それは本当に必要不可欠で、それがないと文の意味記述はできないようなもの?
 - これが本当なら「生命の定義なしに生物学ができない」ってコト
- でも、そんなことはないでしょう??

コトバの使用の基本に戻る

- 事実 A を確認し， B の問いを立てる必要がある
 - A. コトバは “他者操作の道具” である
 - これは単なる(概念)メタファーではなく， ヒトは実際にコトバを使って他者 (しばしば自己も含む) を操作している
 - B. なぜコトバが他者操作にこれほどうまく機能するのか?
- B の解明は言語の認知科学にとって (「普遍文法の探求」 などよりずっと) 重要な課題
 - B の解明は 「言語の進化」 のような(疑似)問題に正しく対処するための前提

コトバの意味の特定の課題

- そのためには,
 - A. 特定のコトバを使うことで実際に何が相手に伝わっているのかを正確に特定
 - B. コトバを使うことでヒトが実際に何をしようとしているのかを正確に特定
- する必要がある
 - A, B は同じではない: 多くの場合 A なしに B は達成不能
 - B は明らかに発話行為論 (Austin, Searle) 的な設定だが, 単に発話行為の分類を与えるだけでは充分ではない

落とし穴

- でも、考えても見よ
 - 特定のコトバを使うことで実際に何が相手に伝わっているのかを正確に特定する
- という目標は,
 - まだゼンゼン達成されていないし,
 - そもそも真剣に取り組まれていないではないか?
- でも、それはなぜ??

浅い理由と深い理由

- その理由は,
 - コトバの意味の (中途半端な) 表示 representations が, 表示すべき対象の詳細な特定 specifications に優先されて来たから
- この「根本問題の先送り」を正当化しているのは
 - 文の意味は, その部分である語句の意味から“構成”されるから, 部分の意味の網羅的記述が得られれば, おのずと文の意味が得られる
- という誤った想定

問題の克服のためには

- 先送りを止めて「部分の意味が全体の意味に依存する」というゲシュタルト質を失わない形で
 - 文 s の意味 $m(s)$ と s を形式的に構成する語句 $W(s) = \{w_1, w_2, \dots, w_n\}$ の意味 $\{m(w_1), m(w_2), \dots, m(w_n)\}$ を
- 同時に記述する手法を開発する必要がある
- 言語使用の現実を支配するのはゲシュタルト性
 - 構成性は (文法を代数系として見なした場合に) 極めて魅力的な性質だが、これは不可避免的にゲシュタルト質の過小評価につながる

要件

- 文レベルの意味を直接記述する必要がある
 - 計算機用の辞書/シソーラスの構築は、環境情報を明示した語句の意味記述を元にして始められるべき
 - さもないと、辞書に書かれている語句の意味と、文を通じて実際に理解される、ゲシュタルト性をもった語句の意味の食い違いが埋められない
- 構成性を使った一見するとカッコいい「説明」より、ベタで退屈だが、本当に役に立つ「記述」を

複層意味フレーム分析

- 以上の要件を満足する技法として提案されたのが複層意味フレーム分析 (MSFA)
 - これは Berkeley FrameNet (BFN: Fillmore, et al. 2003) にヒントを得ているが BFN とは独自に定義された枠組み (Japanese FrameNet (JFN): Ohara et al. 2003)
- ただし,
 - MSFA という名称は (今となっては) どうでもいい
 - 記述単位が本当にフレームと呼ばれるべきなのかどうか (今となっては) どうでもいい

複層意味フレーム分析の 詳細

粒度が細かすぎる??

- MSFA の記述は「一般性を欠く」とか「細かすぎる」とか批判したい人への回答
 - 細かすぎると言うのは簡単だが、文理解の際に達成される理解内容の妥当な記述に、どれほどの粒度の記述が必要なのか、誰もマジメに調べていないではないか!
- 内容理解が達成されるのは、具体的、個別的な状況レベルではないのか?
 - 具体例から一般例に遡ることは可能だが、その逆はムリ
 - 情報の廃棄は容易、喪失情報の復元は至難

ヒトの理解の性質

- ヒトの (文) 理解の状況基盤である
 - 詳細は黒田・中本・野澤 (2005), 中本・黒田・野澤 (2005) を
- このことの反映として, ヒトの(文)理解は意味型のレベルではなく, 意味役割レベルで達成される
 - 従来の意味型ベースの意味記述の限界が帰結
- Berkeley FrameNet (Fillmore et al. 2003; Baker et al. 98) の示唆が正しければ, 理解の単位である状況を “フレーム” という単位で取り出せる (ハズ)

MSFA の実例

1. 大型の台風が九州を襲った.
2. ライオンがインパラの群れを襲った
3. 覆面の男が銀行を襲った.

MSFA の実例

1. 大型の台風が九州を襲った.
2. ライオンがインパラの群れを襲った
3. 覆面の男が銀行を襲った.

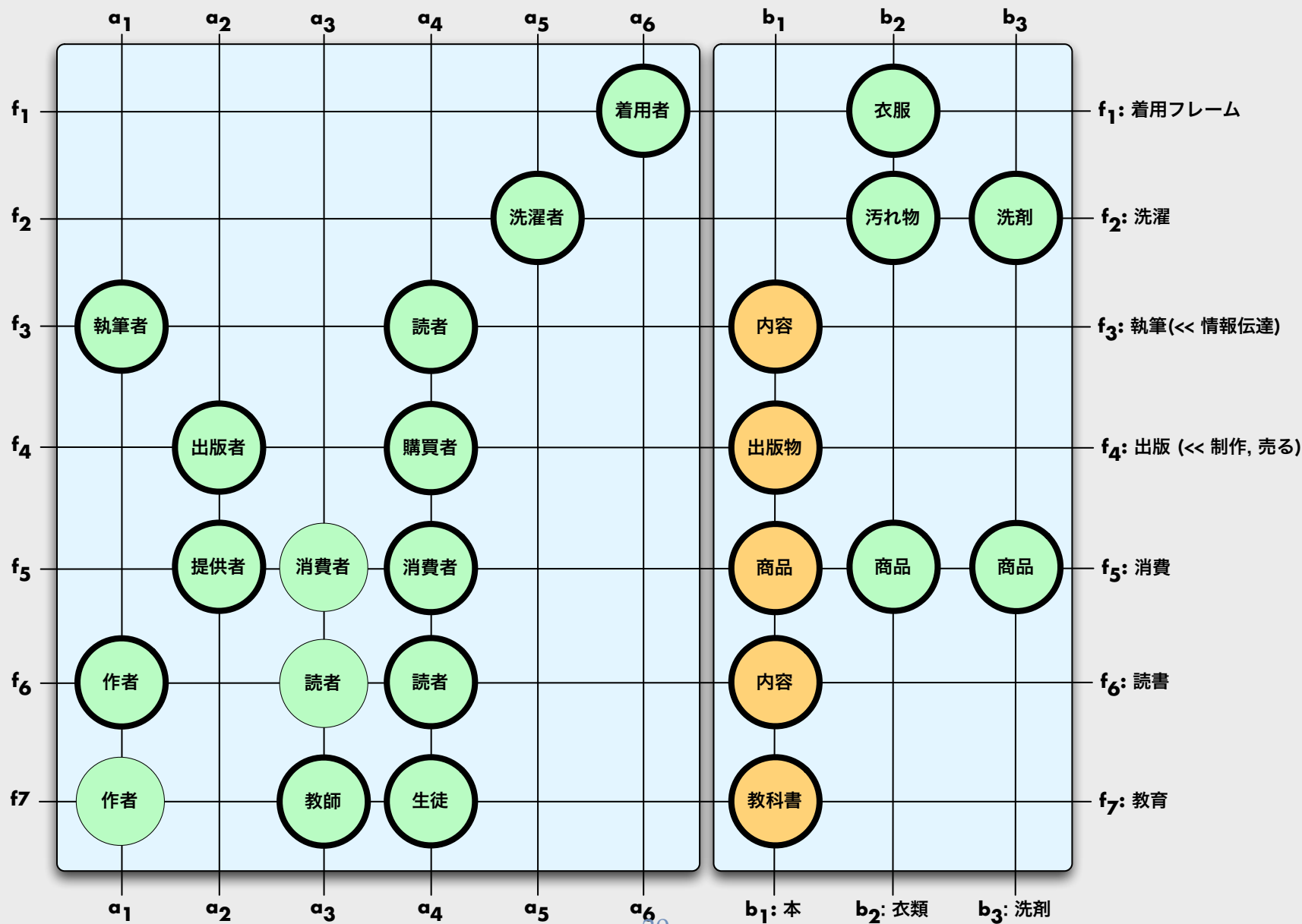
(1) の MSFA

Frame ID	F1	F2	F3	F4	G1	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
Frame-to-Frame Relation		elaborates F3,F4	elaborates F4; elaborates G1	constitutes F3		elaborates F9; targets F6	constitutes F7; elaborates F9	presupposes F6; elaborates F8	elaborates F9	constitutes F8	elaborates F11			elaborates F14		realizes F14; presupposes F14
Frame Name	一時的発生	経路移動	移動	経路	非生命体の活動	攻撃 [+metaphoric]	加害	被災	経験	影響	ヒトの生活	生物の棲息	収容 [+metaphoric]	*構成*	*国土*	区別
*	GOVERNOR				活動内容1											
大型の	規模					大きさ?	規模?	規模?	規模?	規模?						
台風	発生体: EVOKER	移動体: EVOKER	移動体		活動体	攻撃者: EVOKER	加害体: EVOKER	原因	内容	影響体						
*		余波														
が						MARKER	MARKER									
		GOVERNOR			活動内容2											
*	発生地	起点	起点	起点											国外	対象1
*		着点	着点	着点												
*		通過点	経過点	経過点		獲物1	被害体1	被災体1						全体[=日]	国内	対象2
九州						獲物2	被害体2	被災体2	場所		生活地域: EVOKER	棲息地域	収容体	部分: EVOKER		
(の											MARKER		収容物			
人々)						獲物3	被害体3	被災体3	経験者		生活者	生物				
*						獲物4	被害体4	被災体4			住居: EVOKER	住みか				
を						MARKER	MARKER									
襲つ						EVOKER & GOVERNOR	EVOKER & GOVERNOR	様態?								
た						EXTENDER	EXTENDER									
.																

名詞の担う意味役割

- フレームごとに「台風」は次の意味役割を実現
 - <(自然)発生>フレームの<(自然)発生体>; <(経路)移動体>フレームの<(経路)移動体>; <攻撃>[+metaphoric]フレームの<攻撃者>[+metaphoric]; <加害フレーム>の<加害体>; <被害>フレームの<被害の原因>
- これらは動詞 (e.g., 「襲う」) によって付与されるものではない
 - 動詞は単に潜在的な意味役割の集合から幾つかを選択するだけ [cf. *Qualia Structure* Pustejovsky (1991, 1995)]

意味役割の多次元性

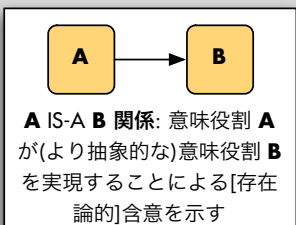
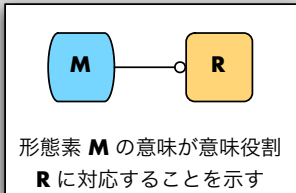
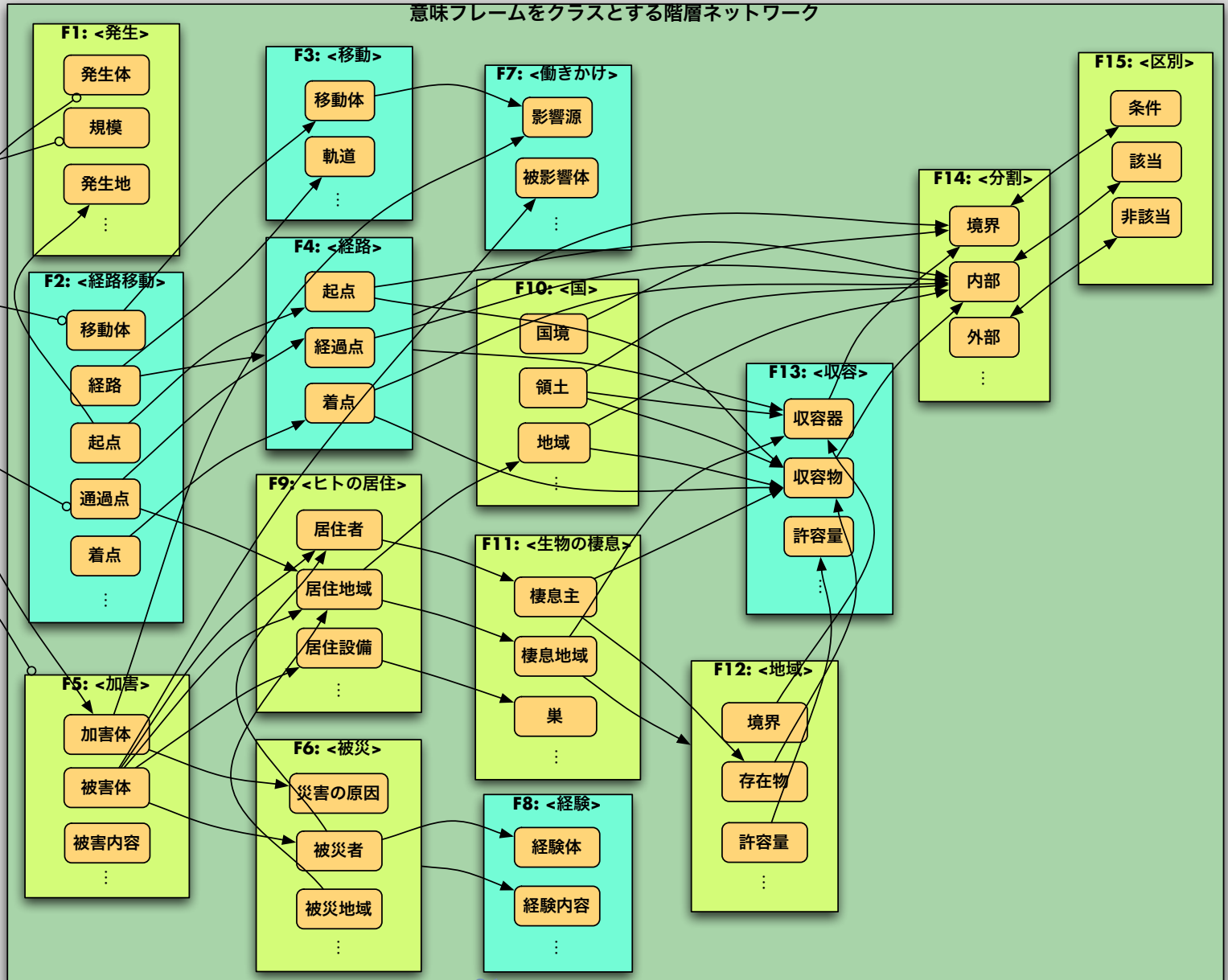


(1) の SFNA

意味フレームをクラスとする階層ネットワーク

自然言語文

大型
の
台風
が
九州
を
襲っ
た



MSFA と SFNA の関係

- MSFA は Semantic Frame Network Analysis (SFNA) への橋渡しの存在
- SFNA は
 - 語による状況の喚起のネットワークの形で表わしたものの
 - 事態のオントロジー階層を表わしたものの
 - 名詞のオントロジー階層を表わしたものの
- 以上のことから
 - シソーラスの構築に深く関係

意味役割名を使った シソーラスの再組織化

意味役割(名)の恩恵

- 意味役割(名) ≠ 意味型(名)
 - 意味型は自然分類を与えるが、意味役割は機能/用途分類を与える
- 従来のシソーラスでは意味型と意味役割名が区別されておらず、混乱が見られる
 - STN: {秋田犬, シェパード(犬), ...}, {岩石, 鉱石, ...}
 - SRN: {盲導犬, 番犬, ...}, {宝石, 敷石, ...}
 - 詳細は黒田・井佐原 (2005), 黒田・中本・野澤 (2005) を
- 意味役割名の一般理論があれば、再組織化可能?

重要な点

- ヒトの理解は意味型のレベルではなく，意味役割レベルで達成される
 - これは，ヒトの理解の状況基盤性の反映
- これが正しいとすれば，意味型分類を基本原則として構築された従来のシソーラス (日本語語彙大系) や，それを元にした意味記述の有用性には限界がある

フレーム喚起の実例

- 京大コーパス S-ID: 950107210-002
 - アルゼンチンの元サッカー選手，ディエゴ・マラドーナ氏が，六日，同国の検察当局に一時身柄を拘束された
 - “元~”は<<引退者: x>が, {<活動: y>から; <役職: z>を}, <理由: w>で, <GOV: 引退>する>フレームを喚起
 - <引退>は<<活動者: x>が, <分野: y>で, <役職: z>として, <GOV: 活動>する>フレームを(前提の関係により)喚起
 - ディエゴ・マラドーナが<引退>と<活動>フレームの x
 - サッカー選手が<引退>フレームの y, z と<活動>フレームの z を実現

典型値効果と命名効果

- **典型値効果**: 対象名 n は, フレーム f を構成する意味役割 $f.r$ の典型値であることによって f を喚起
 - サメ --> <人食い>, イルカ --> <水族館のショーの主演>
- **命名効果**: 役割名 n^* は, フレーム f を構成する意味役割 $f.r$ の名称であることによって f を喚起
 - **役職名**: 大統領, 首相, 社長, 奴隷, 僕 (しもべ),
 - **職業名**: 運転手, 医者, 妻, 夫
 - **状況特有の(意味)役割名**: 運転者, 犯人, 加害者, 被害者, 犠牲者, 被災者

意味役割名の重要性

- 名詞の一部は状況 (= フレーム) に固有の意味役割 (= フレーム要素) の名称になっている
 - 被害者, 犠牲者 → <被害>フレームの構成要素
 - 犯人 → <犯罪>フレームの構成要素
 - 裁判官, 被告(人), 検察(側), 弁護人 → <裁判>フレームの構成要素

まとめ

MSFA は

- 文の意味とゲシュタルト質を保存したまま語句の意味を記述するための手法
- 言語表現と世界知識の結びつきを網羅的に記述するためのコーディング技術
 - 従って、文の意味の「説明」理論ではない
- 言語学者の意味分析の (得てして自己目的的で、盲目的な) 仕事を認知科学、自然言語処理へ橋渡し

課題, 残された問題

- 作業規模の拡大, 記述量の増加
 - Open Development にするのが理想でしょうね
- 記述内容の質の向上
 - 記述内容の標準化! 標準化!! 標準化!?!
 - まずは作業者の結果の評価方法の確立から
 - 記述内容の国際化???
 - これは難しい課題です. 上位オントロジーなら, ともなく下位オントロジーの外国語への翻訳は大変

謝辞

以下の方々にお世話になりました

内山 将夫 (NICT)

金丸 敏幸 (京都大学 大学院)

李 在鎬 (NICT)

中本 敬子 (京都大学教育学研究科)

野澤 元 (NICT)

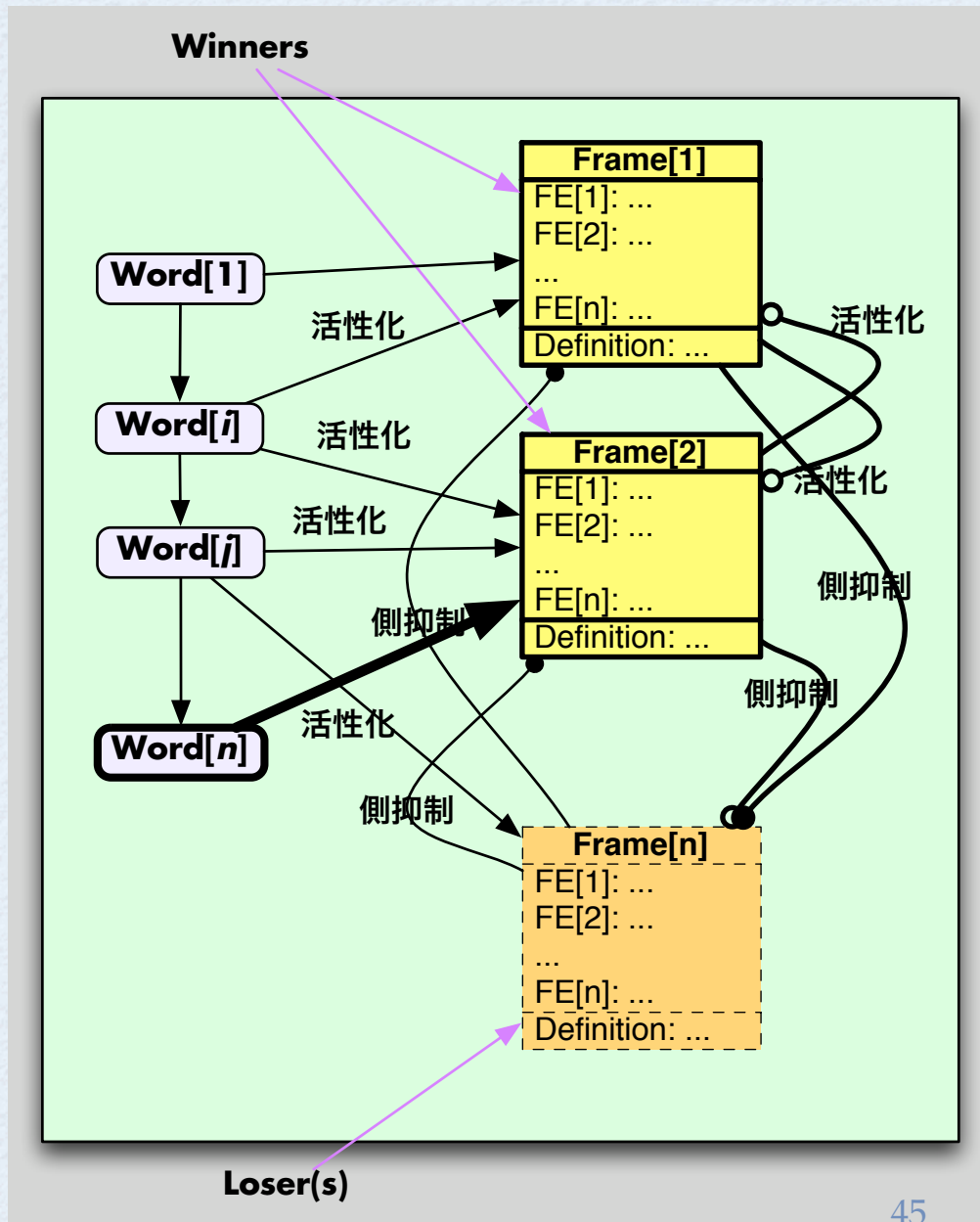
MSFA の設計思想

付録

フレーム喚起の仕組み

- 語 w は、何らかの(意味)フレーム f の構成要素 $f.r$ (= 意味役割) の喚起体 evoker であり、 w が使用されると少なくとも一つの f が喚起され
 - 一般に w_i が喚起するのはフレーム群 $F_i = \{f_{i,1}, f_{i,2}, \dots, f_{i,n}\}$
- 文 $s = w_1 w_2 \dots w_n$ の意味は w_i がおのこの喚起するフレーム群 $\{F_1, F_2, \dots, F_n\}$ の統合として与えられる
 - フレームの統合はフレームの活性化の伝播と側抑制の結果だと考える

フレーム単位の語義の選択



- 語 w の意味の選択の単位はフレーム = 状況
 - フレームの選択は自然選択に似たプロセス
- 選択の単位がフレームだということは“全体の意味が部分の意味に優先する”ということ
 - 強い意味での構成性は前提にできない

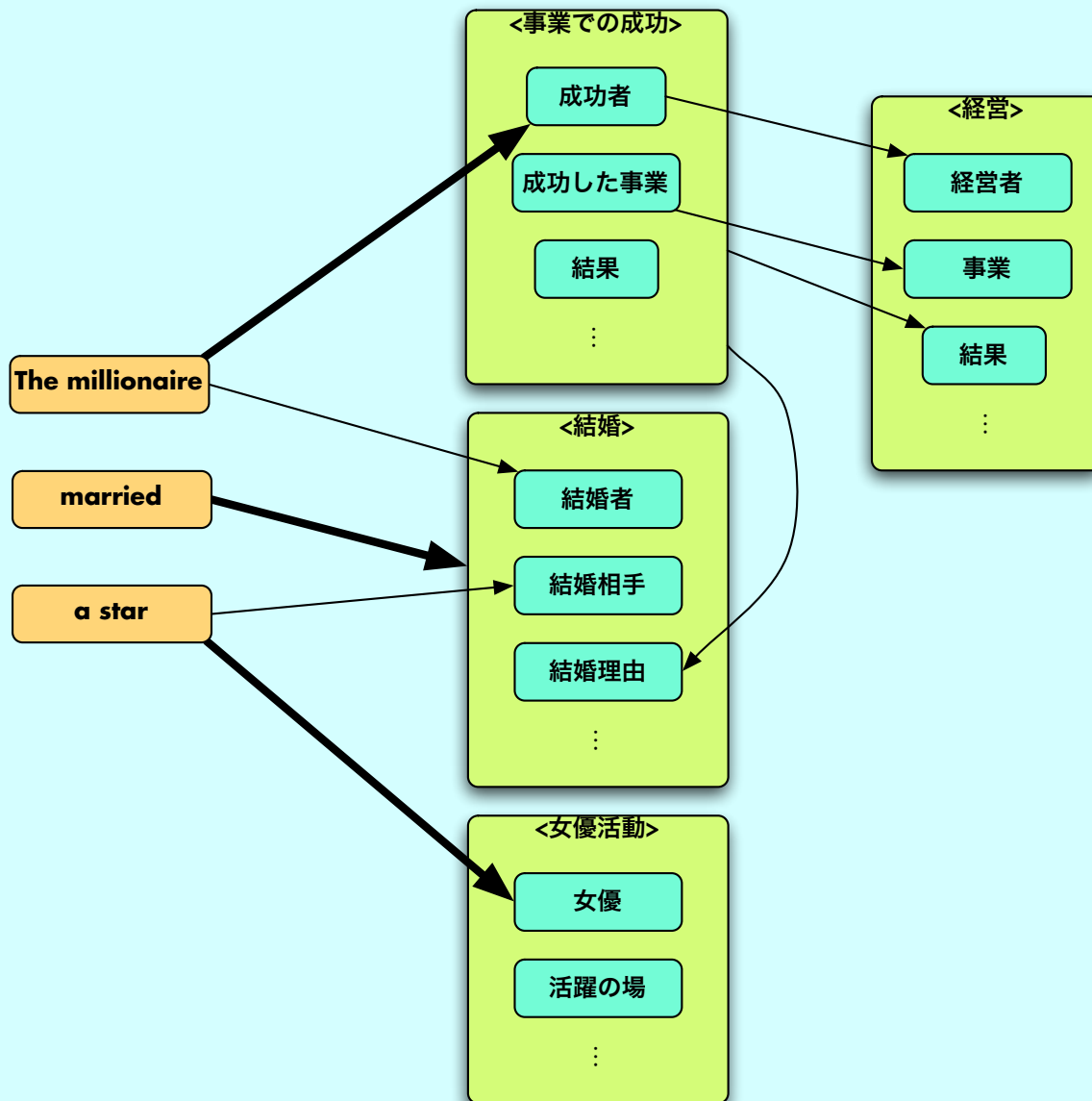
文意の並列分散性

- 文を構成する語のおのおのが互いに独立に，同時多発的に意味フレームを喚起し，
 - 語による状況の喚起は“有か無か”の二値 $\{1, 0\}$ でなく $[0, 1]$ の連続値の重み付きの結合 (ただし適当な判別関数が必要)
- このようにして喚起されたフレーム群の“統合”によって文意が構成される

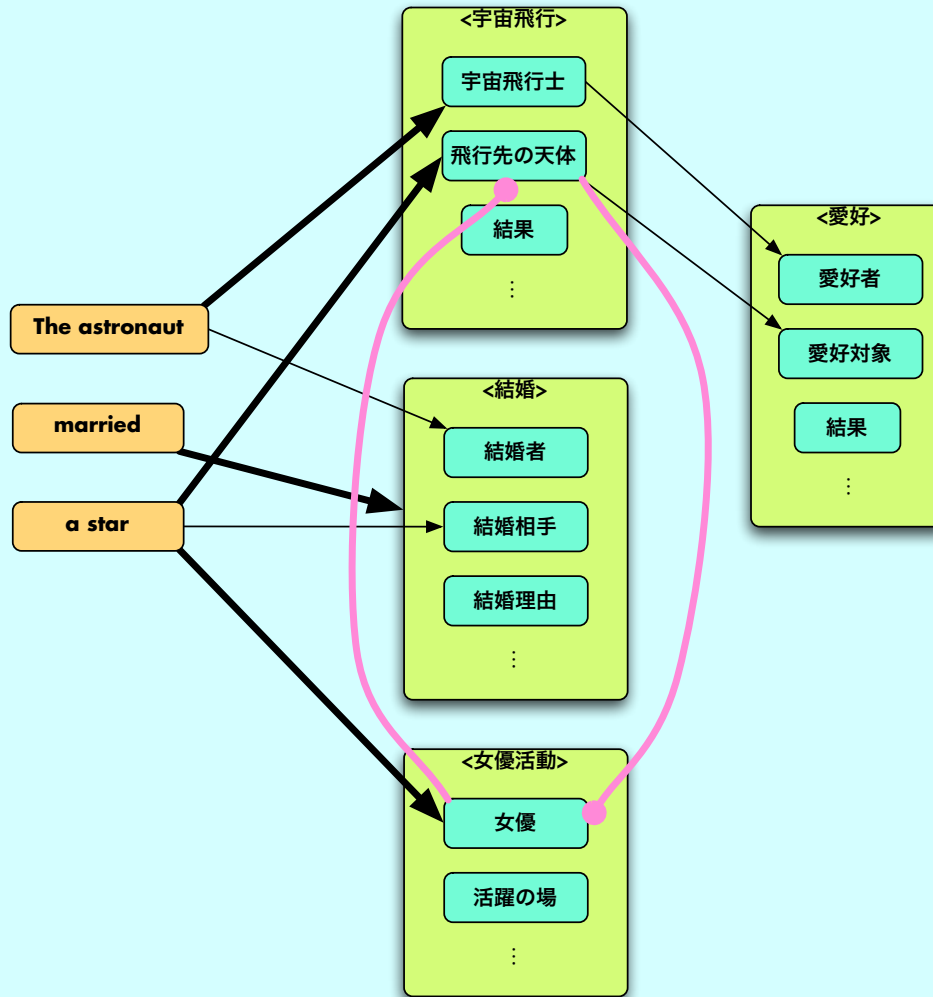
フレーム統合の実例

- Minsky (1986) の古典的な例を取り上げる
 - The millionaire married a star.
 - The astronaut married a star.

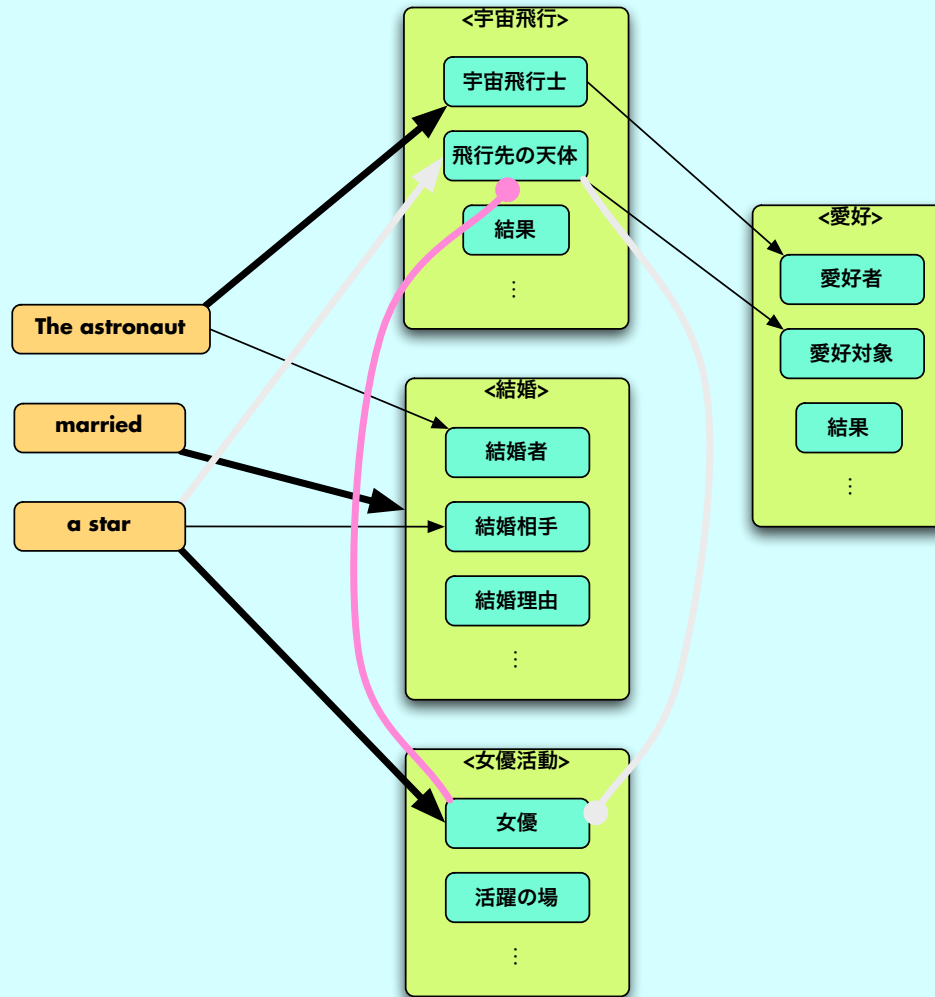
簡単な実例 [1]



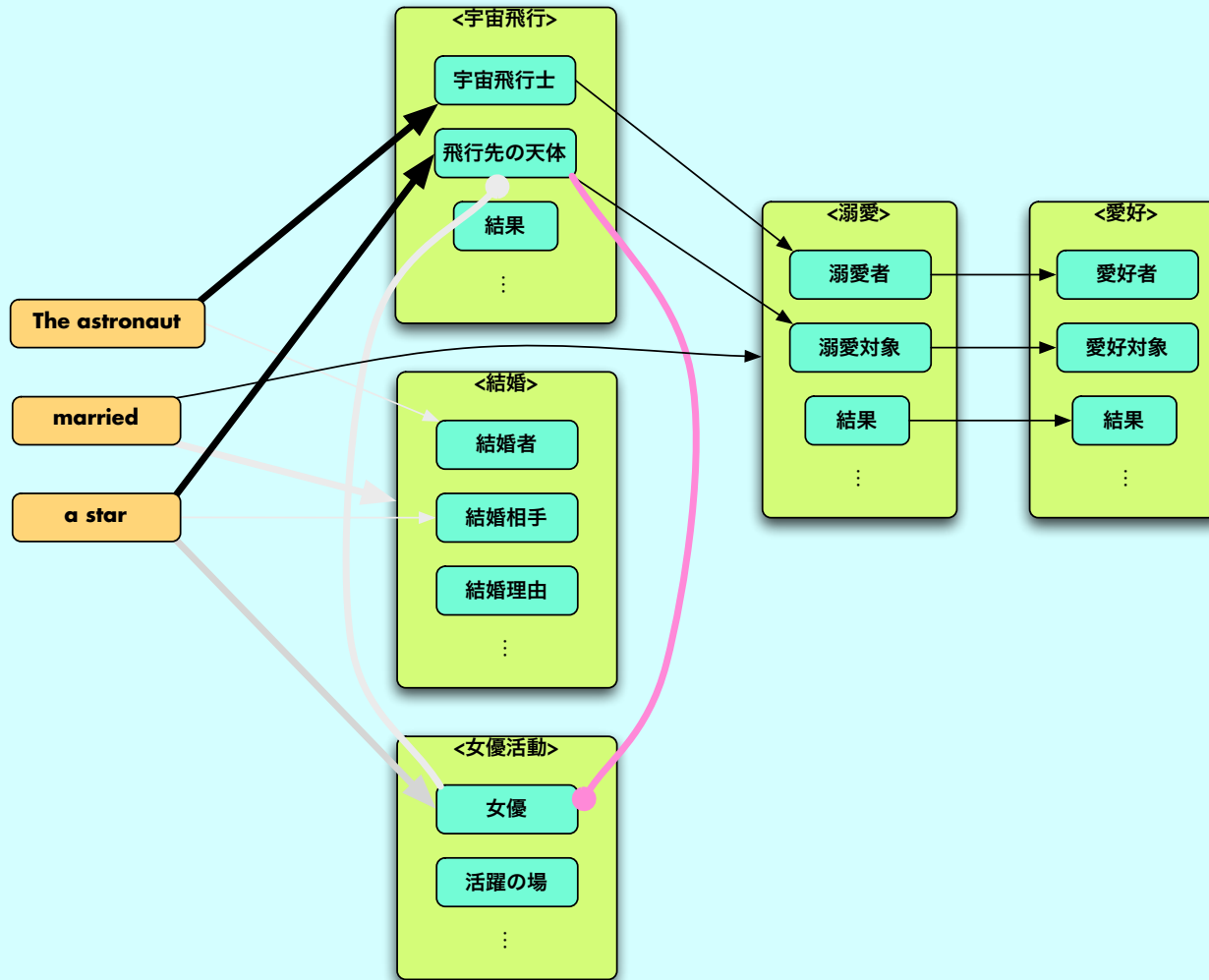
簡単な実例 [2]



矛盾の解消 [1]



矛盾の解消 [2]



フレームへの引込み効果

- (1) の (a) star には対象名と意味役割名の曖昧性がなく, (2) の (a) star には対象名と意味役割名の曖昧性がある
 1. The millionaire married a star.
 2. The astronaut married a star.
 3. The man married a star.
- (1), (2) に [(3) に対して] プライミング効果があるというだけでは不十分

重要なこと

- Minsky (1986) のモデル (cf. Schank and Abelson 1978) は基本的に妥当だが、重要なのは
 - どの語がどのフレームに（どれぐらいの強さで）結びついているか、あるいはどのフレームを（どれぐらいの強さで）喚起するかを網羅的に調べること
 - これが明示されない限り「語がフレームを喚起し、その結果として文意が決定される」という定式化は（仮に正しくても）応用的価値はない
- Scalability の問題に挑戦する必要あり