

屈折パラダイムの(心内)表現はどうなっているのか？

How are inflectional paradigms represented (in the mind)?

FCA meets Czech declension

Kow Kuroda/黒田 航

Kyorin University/杏林大学

22nd Annual Meeting of the NLP Association, Japan, at Tohoku University,
2016-03-10 (Thu)

事の始まり

- チェコ語を2013年に始めて曲用 (declension) に悩まされている
 - この複雑性の正体は何なのだ！？
 - 例えば Czech-English English-Dictionary (Fronek 2010) には、名詞について全部で68種類(男性27種類, 中性12種類, 女性15種類), 形容詞について全部で4種類の曲用クラスが記載されているが, これらの相互関係はどうなっているのか？
 - もっと効率の良い覚え方はないものか！？
- という (苛立ち混じりの) 疑問の解消が目的
 - 言語学を学び始めた頃に読んだ形態論の入門書 [Carstairs-McCarthy 1992] に, 形態論で paradigm の研究は不十分だと書いてあったし

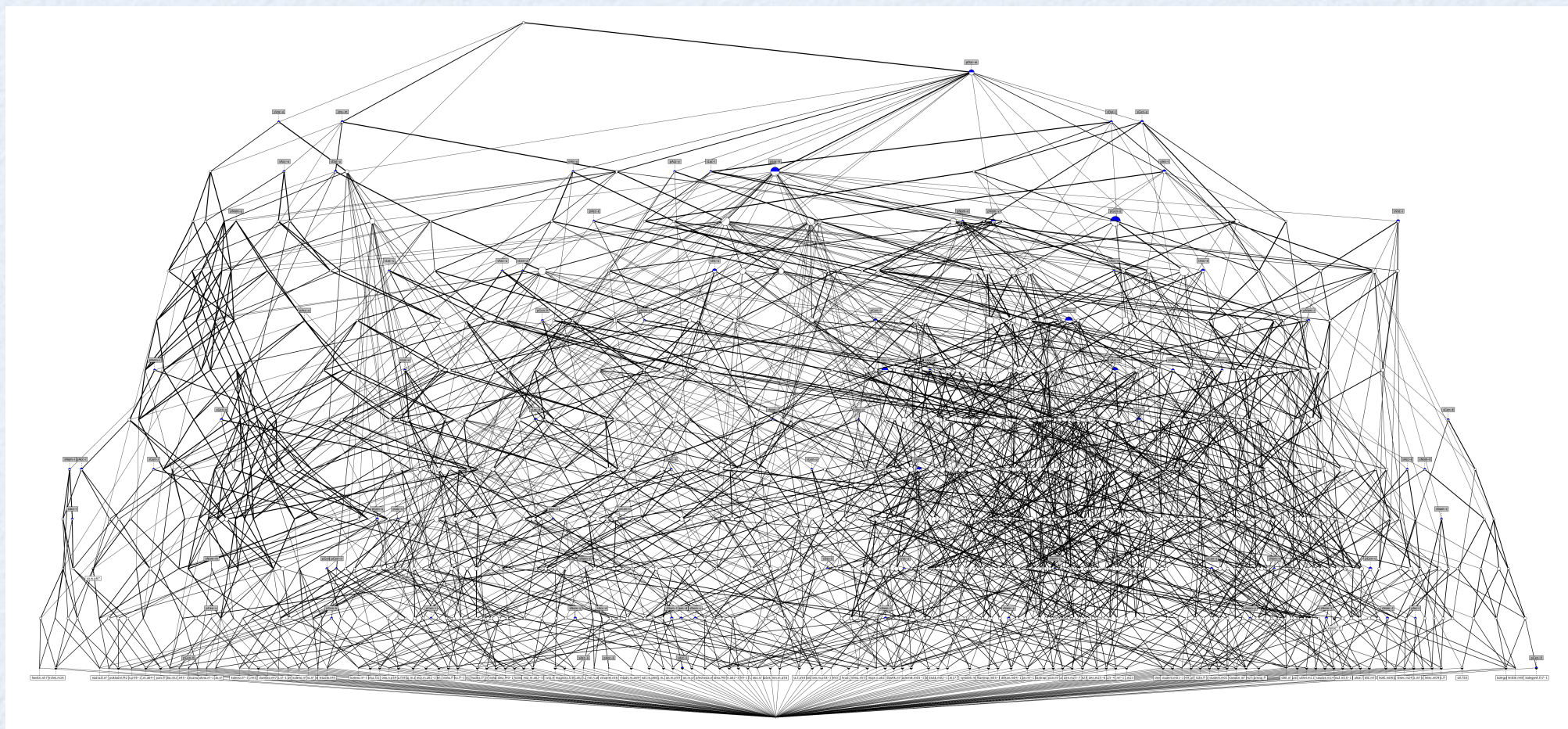
簡単な例

	Masc 01	Masc 11	Neuter 44	Neuter 53	Fem. 39	Fem. 36	Ending types
Nominative	student_	učitel_	místo	album	škola	ulice	_, o, um, a, e
Genitive	studenta	učitele	místa	alba	školy	ulice	a, e, y
Dative	studentu	učiteli	místu	albu	škole	ulici	u, i, e
Accusative	studenta	učitele	místo	album	školu	ulici	a, e, o, um, u, i
Vocative	studente	učiteli	místo	album	školo	ulice	e, i, o, um, o, e
Locative	studentu	učiteli	místě	albu	škole	ulici	u, i, ě, u, e, i
Instrumental	studentem	učitelem	místem	albem	školou	ulicí	i, é, a, y, í
Nominative	studenti	učitelé	místa	alba	školy	ulice	i, é, a, y, e
Genitive	studentů	učitelů	míst_	alb_	škol_	ulic_	ů, _
Dative	studentům	učitelům	místům	albům	školám	ulicím	ům, ám, ím
Accusative	studenty	učitele	místa	alba	školy	ulice	y, e, a, e
Vocative	studenti	učitelé	místa	alba	školy	ulice	i, é, y, e
Locative	studentech	učitelích	místech	albech	školách	ulicích	ech, ích, ách
Instrumental	studenty	učiteli	místy	alby	školami	ulicemi	y, i, ami, emi

例からわかること

- form-to-function では語形の polymorphism か派生語尾の operator-overloading が生じている
 - この現象を言語学/形態論では **syncretism** という
- 語尾でパラダイム中の位置を判別するのは (一定の傾向があるとは言え) 無理
 - 名詞の性別に頼るのは, それが主格の語尾がわからないと判定できないため, 問題先送りにしかない
 - 全部で14個の場合があるのに, 基本的に母音 (a, e, é, i, y, o, u, û) でマークしようとする傾向があり, function-to-form の写像で overload が起きている

参考データ: 語尾を属性にを使ったFCA (属性と対象の削減後)



問題

- 語尾が曲用クラス識別キーにならないなら、どんな情報が識別キーになっているのか？
- 曲用クラス同士の類似度は測れないのか？

検証する仮説 1/2

- 仮説 H
 - チェコ語の名詞/形容詞の 14個の語形の(非)同一性の集合が曲用クラスの識別キーなのでは？
- 関連研究 (paradigm economy, inflection classes)
 - [Ackerman, *et al.*, 2009, Ackerman & Malouf, 2013; Carstairs-McCarthy, 1992, 1994; Blevins, 2004, 2013; Wurzel, 1987]

検証する仮説 2/2

- 仮説 Hの含意
 - 名詞 N の曲用の個々の語形は，曲用クラス (i.e., パラダイム) の認識に間接的にしか係わらない
 - 個々の語の終り方は名詞の性別に関係するだけ
 - パラダイムは表層語形に対して二次の一般化 (2nd order generalization)

課題の定義

● 方針

- H が正しいなら曲用クラスの教師なし分類が可能である
 - 発表後の訂正
 - 最適化の手順が教師信号になっているので、教師なしは誤り
 - せいぜい半教師あり分類
 - この分類を Formal Concept Analysis (FCA) [Ganter & Wille, 1999; Ganter, et al. (eds.), 2005; 鈴木・室伏 2007] を使って検証する

● 実装

- Pairwise identity-network (PIN) (後述) を FCA に与える形式文脈の属性に使う

この後の概要

- ~~問題の定義~~
- 分析用データ D の構築
- FCAの概要
- データ D へのFCAの適用結果と考察
- まとめと今後の課題


```
Voc, pGen=pLoc, pGen=pIns, pDat=pAcc, pDat=pVoc, pDat=pLoc, pDat=pIns, pAcc=pVoc, pAcc=pLoc,  
pLoc, pVoc=pIns, pLoc=pIns, masc, fem, neut-  
,  
0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
-1 ,  
0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
f.a69 ,  
1,0,0,0,0,0,0,0,0,0,0,0,1,0,0,1,0,1,0,0,1,1,0,0,0,0,0,0,1,0,1,0,0,1,1,0,0,0,0,0,0,0,0,0,  
,0,0,0,1,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
n.a69 ,  
1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
m.a70 ,  
1,0,0,1,0,0,1,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
,0,0,1,0,0,1,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
-1
```

分析用データの開発

データの準備

- 次の資料に記載されている名詞と形容詞の曲用を手動で電子データ化
 - 『チェコ語基本1500語』 (金指 1998)
 - *Czech-English English-Czech (Compact) Dictionary* (Fronek 2010)
 - `_.m09`, `_.f39`, `_.46` のような分類コードは Fronek (2010) を適用
 - 金指 (1998) から採取したデータは `_.m?`, `_.f?`, `_.n?` と表記
- その後、自作の Python スクリプトでPINの 0/1 ベクトルに変換
 - これが Concept Explorer 1.3 に与える formal context になる

元データの様子

```
1 % attribute groups are separated by ";".-
2 % gender value is separated by ":".-
3 # sNom, sGen, sDat, sAcc, sVoc, sLoc, sIns; pNom, pGen, pDat, pAcc, pVoc, pLoc, pIns-
4 -
5 %% from Kanazashi's database, unlisted in Fronek-
6 m?: pán, pána, pánovi, pána, páne, pánovi, pánem; pánové, pánů, pánům, pány, pánové, pánech, pány-
7 m?: pán, pána, pánu, pána, páne, pánu, pánem; páni, pánů, pánům, pány, páni, pánech, pány-
8 m?: průvodce, průvodce, průvodci, průvodce, průvodce, průvodci, průvodcem; průvodcové, průvodců,
... průvodcům, průvodce, průvodcové, průvodcích, průvodci-
9 m?: předseda, předsedy, předsedovi, předsedu, předsedo, předsedovi, předsedou; předsedové,
... předsedů, předsedům, předsedy, předsedové, předsedech, předsedy-
178... mladých, mladými-
179 f.a69: mladá, mladá, mladá, mladá, mladá, mladá, mladá; mladá, mladých, mladým, mladá, mladá,
... mladých, mladými-
180 n.a69: mladá, mladého, mladému, mladá, mladá, mladém, mladým; mladá, mladých, mladým, mladá, mladá,
... mladých, mladými-
181 m.a70: jarní, jarního, jarnímu, jarního, jarní, jarním, jarním; jarní, jarních, jarním, jarní,
... jarní, jarních, jarními-
182 -
183 %% end of file-
184
```


Pairwise Identity-network (PIN)

Table 6は論文に非掲載

Table 3: Matrix specifying 91 pairwise formal identities

	sGen	sDat	sAcc	sVoc	sLoc	sIns	pNom	pGen	pDat	pAcc	pVoc	pLoc	pIns
sNom	sN=sG	sN=sD	sN=sA	sN=sV	sN=sL	sN=sI	sN=pN	sN=pG	sN=pD	sN=pA	sN=pV	sN=pL	sN=pI
sGen	-	sG=sD	sG=sA	sG=sV	sG=sL	sG=sI	sG=pN	sG=pG	sG=pD	sG=pA	sG=pV	sG=pL	sG=pI
sDat	-	-	sD=sA	sD=sV	sD=sL	sD=sI	sD=pN	sD=pG	sD=pD	sD=pA	sD=pV	sD=pL	sD=pI
sAcc	-	-	-	sA=sV	sA=sL	sA=sI	sA=pN	sA=pG	sA=pD	sA=pA	sA=pV	sA=pL	sA=pI
sVoc	-	-	-	-	sV=sL	sV=sI	sV=pN	sV=pG	sV=pD	sV=pA	sV=pV	sV=pL	sV=pI
sLoc	-	-	-	-	-	sL=sI	sL=pN	sL=pG	sL=pD	sL=pA	sL=pV	sL=pL	sL=pI
sIns	-	-	-	-	-	-	sI=pN	sI=pG	sI=pD	sI=pA	sI=pV	sI=pL	sI=pI
pNom	-	-	-	-	-	-	-	pN=pG	pN=pD	pN=pA	pN=pV	pN=pL	sN=pI
pGen	-	-	-	-	-	-	-	-	pG=pD	pG=pA	pG=pV	pG=pL	pG=pI
pDat	-	-	-	-	-	-	-	-	-	pD=pA	pD=pV	pD=pL	pD=pI
pAcc	-	-	-	-	-	-	-	-	-	-	pA=pV	pA=pL	pA=pI
pVoc	-	-	-	-	-	-	-	-	-	-	-	pV=pL	pV=pI
pLoc	-	-	-	-	-	-	-	-	-	-	-	-	pL=pI

Table 4: Identify matrix of *okno*

0	0	1	1	0	0	0	0	0	0	0	0	0	0
-	0	0	0	0	0	1	0	0	1	1	0	0	0
-	-	0	0	0	0	0	0	0	0	0	0	0	0
-	-	-	1	0	0	0	0	0	0	0	0	0	0
-	-	-	-	0	0	0	0	0	0	0	0	0	0
-	-	-	-	-	0	0	0	0	0	0	0	0	0
-	-	-	-	-	-	0	0	0	0	0	0	0	0
-	-	-	-	-	-	-	0	0	1	1	0	0	0
-	-	-	-	-	-	-	-	0	0	0	0	0	0
-	-	-	-	-	-	-	-	-	0	0	0	0	0
-	-	-	-	-	-	-	-	-	-	1	0	0	0
-	-	-	-	-	-	-	-	-	-	-	0	0	0
-	-	-	-	-	-	-	-	-	-	-	-	1	0
-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	0

Table 6: Identify matrix of *okno*

0	0	1	1	0	0	0	0	0	0	0	0	0	0
-	0	0	0	0	0	1	0	0	0	0	1	1	0
-	-	0	0	0	0	0	0	0	0	0	0	0	0
-	-	-	1	0	0	0	0	0	0	0	0	0	0
-	-	-	-	0	0	0	0	0	0	0	0	0	0
-	-	-	-	-	0	0	0	0	0	0	0	0	0
-	-	-	-	-	-	0	0	0	0	0	0	0	0
-	-	-	-	-	-	-	0	0	0	0	1	1	0
-	-	-	-	-	-	-	-	0	0	0	0	0	0
-	-	-	-	-	-	-	-	-	0	0	0	0	0
-	-	-	-	-	-	-	-	-	-	0	0	0	0
-	-	-	-	-	-	-	-	-	-	-	1	0	0
-	-	-	-	-	-	-	-	-	-	-	-	0	0
-	-	-	-	-	-	-	-	-	-	-	-	-	0

(c) Kow Kuroda: How are inflectional paradigms represented? at NLP22, 2016

検証する仮説の再解釈

- PIN で 0 (=false) が多い曲用ほどパラダイム内部の語形の中和の度合いが低い
 - 曲用のない言語では, PINを構成するのは14個の1 (true) 列
- 最適性理論 (McCarthy 2008; Prince and Smolensky 2004) の観点で考えると, 仮説 H は制約 local anti-syncretism 違反の許容度
- Anti-syncretism (global):
 - For forms F and G , $F \neq G$ if F and G bear distinct functions everywhere.
- Anti-syncretism (local):
 - For forms F and G , $F \neq G$ if F and G bear distinct functions within such paradigm P that $P = \{\dots, F, \dots, G, \dots\}$.

Concept Explorer

Update: Clear dependent

	A	B	C	D	E	F	G	H	I	
		sNom=sAcc	sNom=sVoc	sNom=p...	sGen=sDat	sGen=sAcc	sGen=sVoc	sGen=sIns	sGen=pAcc	sDat=
pán.m?						×				
pán.m?-1						×				
průvodce...	×		×			×	×		×	
předseda...									×	
hrdina.m?									×	
týden.m?	×									
týden.m?-1	×									
přítel.m?				×		×			×	
přítel.m?-1				×		×			×	
klíč.m?	×								×	
tramvaj.f?	×								×	
růže.f?			×				×		×	
noc.f?	×				×		×		×	
věc.f?	×				×		×		×	
paní.f?	×		×	×	×	×	×	×	×	
hudba.f?									×	
město.n?	×		×						×	
město.n?-1	×		×						×	
nádraží.n?	×		×	×	×	×	×		×	
centrum.n?	×		×						×	
okno.n?	×		×						×	
okno.n?-1	×		×						×	
kotě.n?	×		×						×	
dítě.n?	×		×						×	
oko.n?	×		×						×	
ucho.n?	×		×						×	
ruka.f?										
ruka.f?-1										
noha.f?									×	
noha.f?-1									×	

Context Editor

FCAの概要

FCAって何ですか？ 1/3

- 形式概念を<対象集合, 属性集合>の対と定義し, 概念間の関係を束構造 (lattice) で記述

- 初出は Ganter & Wille 99

- 表1.1と図1.3は鈴木・室伏 07から借用

- context の行は対象に, 列は属性=素性に対応 (x=true)

表1.1 コンテキスト表

	a	b	c	d	e
1		×			
2		×	×		×
3	×	×	×		×
4	×		×	×	
5		×	×	×	×

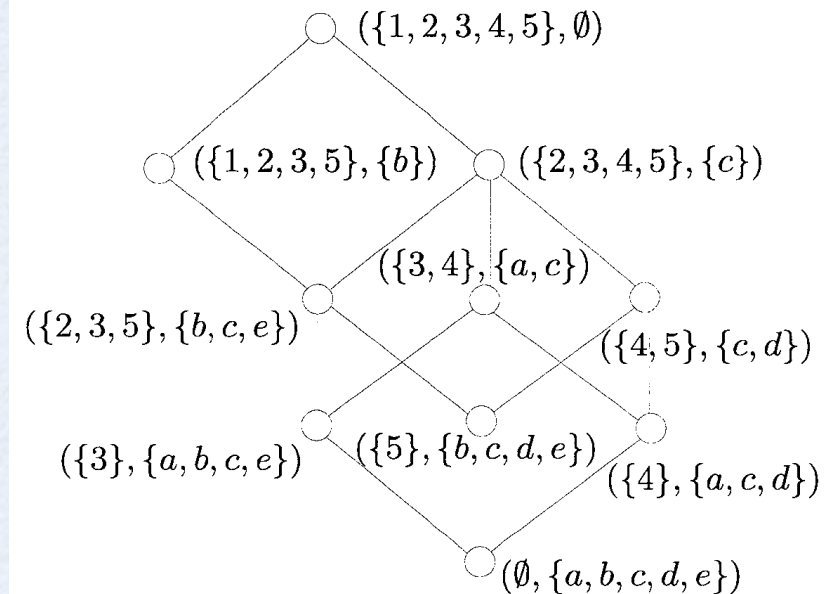


図1.3 コンセプトラティス

FCAって何ですか？ 2/3

表1.4 スポーツのコンテキスト

	a. 球技	b. 道具	c. ネット	d. 陸上	e. 屋内	f. 団体	g. 点数
1. 野球	×	×		×		×	×
2. サッカー	×			×		×	×
3. 相撲				×	×		
4. テニス	×	×	×	×			×
5. バレー	×		×	×	×	×	×
6. ゴルフ	×	×		×			×
7. マラソン				×			
8. 水泳					×		
9. ラグビー	×			×		×	×
10. レスリング				×	×		

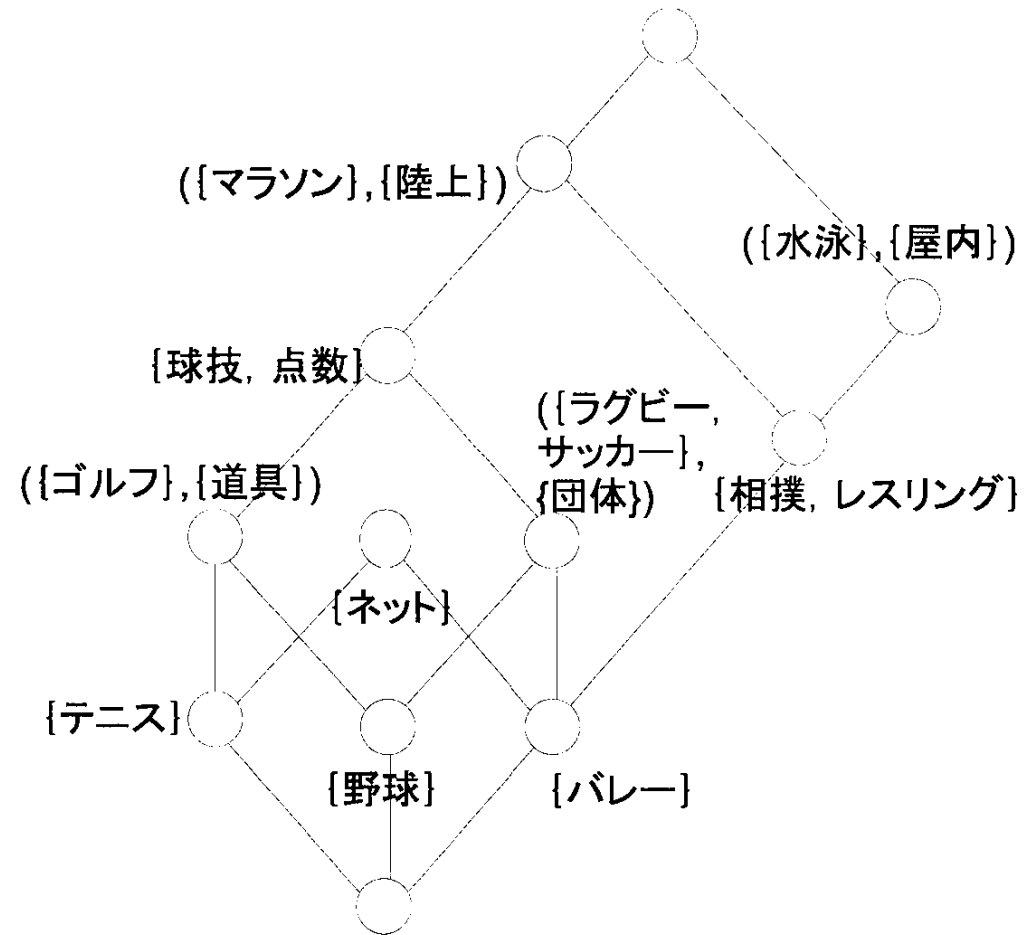
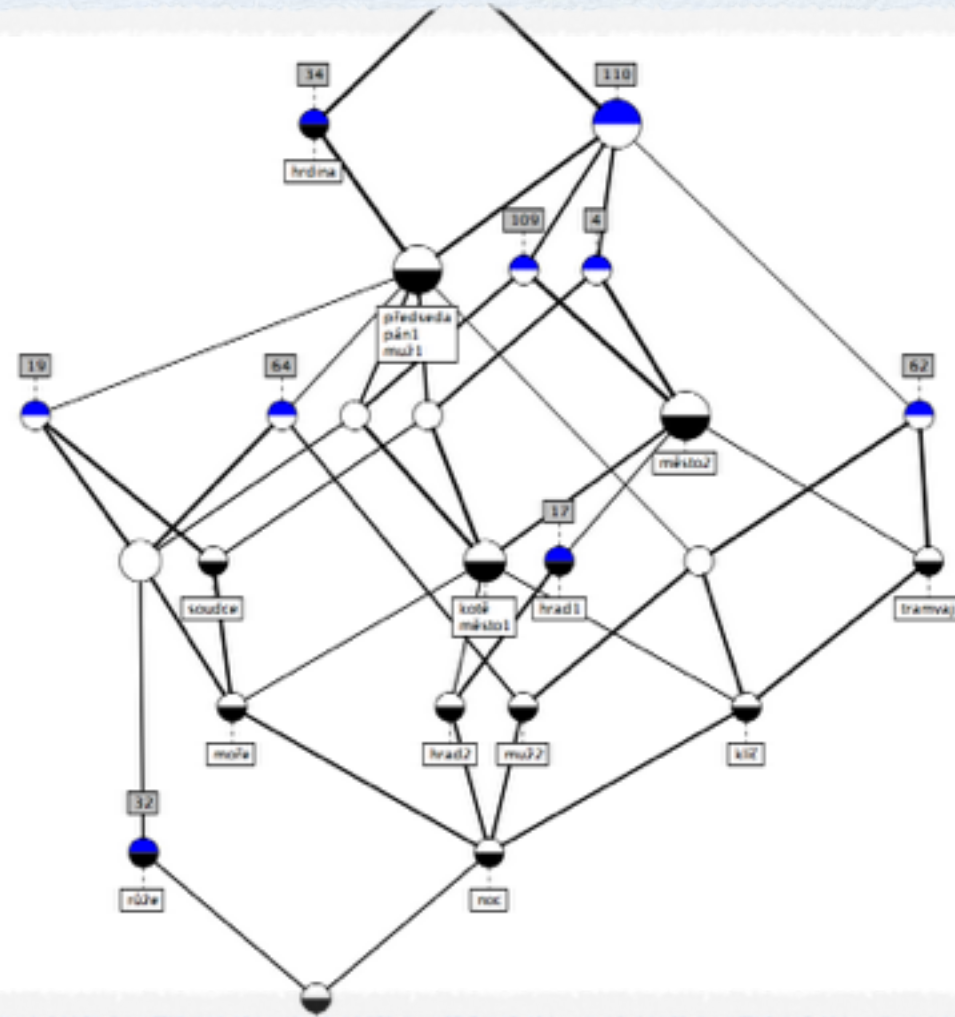


図1.6 スポーツのコンセプトラティス

表1.4と図1.6 は鈴木・室伏 07
から借用

FCAって何ですか？ 3/3

- 利点
 - 体系性を分類束で表現
 - 非排他分類 = soft clustering が可能
 - 分類木を無理やり作らない
- 不利点
- 扱えるのは離散的データのみ
 - 統計処理と相性が良くない
 - R パッケージに連続値の扱える Numerical FCA (nFCA) もあるのですが、本研究で利用した ConceptExplorer ほど使い勝手が良くない



FCAの適用結果

FCA の最適化 1/2

- FCA は全自動で最適な結果を与えない
 - 最適な結果を得るには、機械学習で素性選択に相当する作業が必要
- 最適性の条件
 - 1. 不整合性の回避 (clash が発生: ConceptExplorer では赤いリンク)
 - 2. 過剰分類の回避 (過剰な分類は空所が多く発生)
 - 3. 過少分類の回避 (空所が少なくても分離が悪いのはダメ)
 - 4. 幾何学的安定性の高さ (対象性高いほど高い)
 - これらは互いにトレードオフの関係にある

FCA の最適化 2/2

- 第一段階はFCA/Concept Explorer が実行する自動のデータ圧縮
 - ConExp は Reduce context = Reduce attributes x Reduce objects を実装
- 第二段階は手動で実行
 - Reduce context を実行すると object が不可避免的に削減されるので、Reduce attributes のみを実行する
- 実際には I. Reduce attributes の実行で第一段階を実現，それに II. Reduce objects の追加実行した状態で，第二段階の最適化を行い，目安が付いた段階で II を undo する
 - 第二段階はすべて手動で，非常に時間がかかります

予備調査でわかった事

- 第二段階で相当の試行錯誤した結果，次が判明
 - 名詞と代名詞と形容詞を一緒にすると体系性が下がる
 - 代名詞と形容詞は同類だが，名詞は別
 - 名詞の中でも，女性名詞は別
 - 男性名詞と中性名詞の類似度は高い
 - 特に男性不活動体と中性名詞は実質的に同一クラス

おまけ

- 経験則

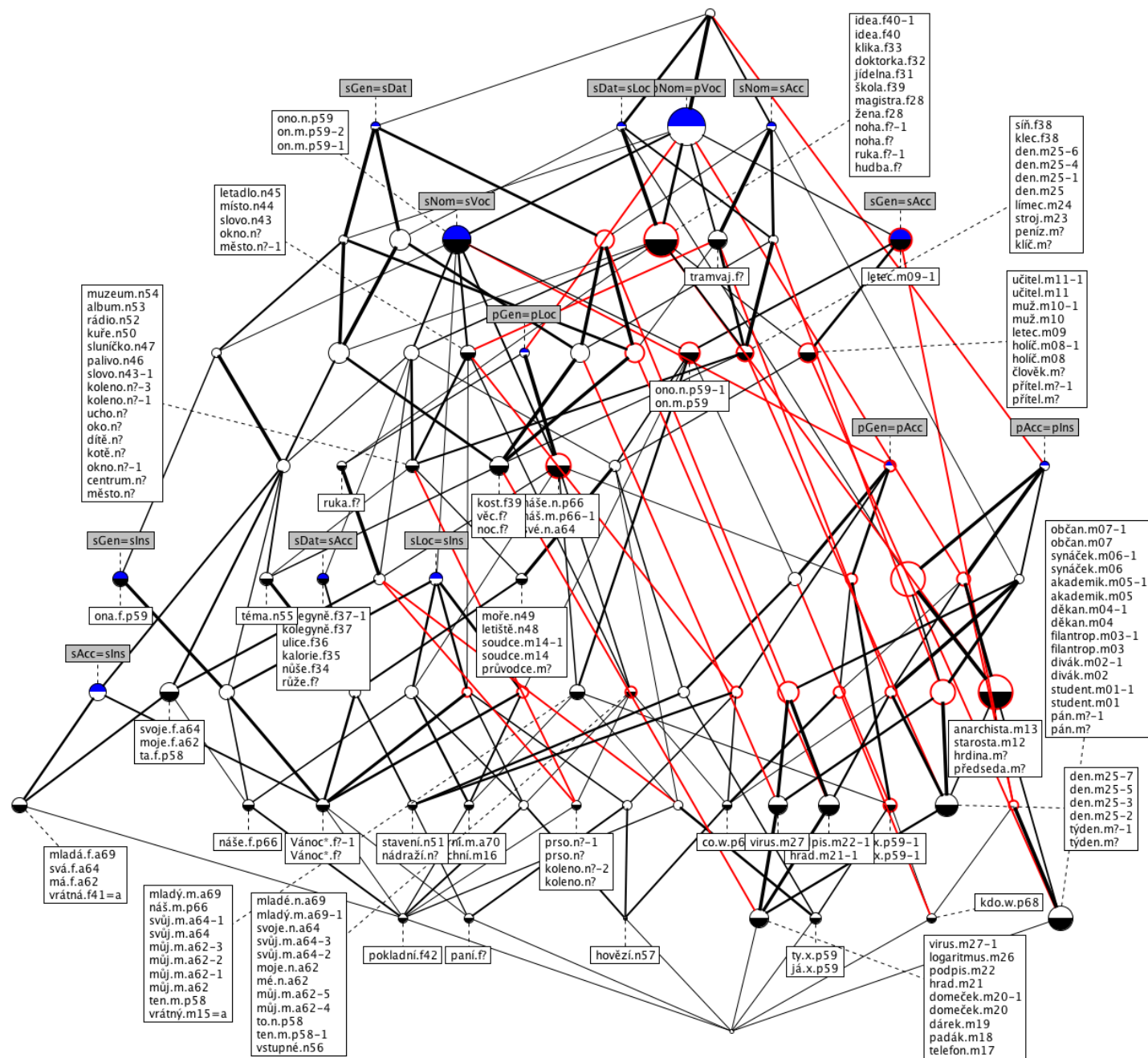
- 奇妙な破綻が生じる場合、土台のエンコーディングが妥当か見直すのは非常に大切

- 理由

- 第二段階の最適化の試行錯誤中、D の構築 (つまりエンコーディング) に間違いがあった事が判明

二段階の最適化の結果

- 論文の Figure 1
- 属性削減ありの全対象の FCA
- 赤いリンクが不整合 = collapse を示す
- 潜在的な対称性は高い



(c) Kow Kuroda: How are inflectional paradigms represented? at NLP22, 2016

自動属性削減で残った29属性

- 91個中の29個
(=0.3186)
 - sNom=sAcc
 - sNom=sVoc
 - sNom=pGen
 - sGen=sDat
 - sGen=sAcc
 - sGen=sVoc
 - sGen=sIns
 - sGen=pAcc
 - sDat=sAcc
 - sDat=sLoc
 - sAcc=sIns
 - sAcc=pNom
 - sAcc=pAcc
 - sVoc=sLoc
 - sVoc=pNom
 - sVoc=pAcc
 - sVoc=pIns
 - sLoc=sIns
 - sLoc=pNom
 - sLoc=pIns
 - sIns=pGen
 - sIns=pDat
 - sIns=pLoc
 - pNom=pVoc
 - pGen=pAcc
 - pGen=pLoc
 - pAcc=pVoc
 - pAcc=pIns
 - pVoc=pIns

属性削減の裏側

- 1 < 15 > sNom=sVoc ==> pNom=pVoc;
- 2 < 4 > sNom=pGen ==> sGen=sAcc sGen=pAcc sDat=sLoc pNom=pVoc;
- 3 < 10 > sGen=sDat ==> pAcc=pVoc;
- 4 < 17 > sGen=sAcc ==> pNom=pVoc;
- 5 < 11 > sGen=sVoc ==> pNom=pVoc;
- 6 < 10 > sGen=sLoc ==> pAcc=pVoc;
- 7 < 21 > sGen=pAcc ==> pNom=pVoc;
- 8 < 7 > sDat=pIns ==> sGen=pAcc pNom=pVoc;
- 9 < 5 > sAcc=sLoc ==> sNom=sVoc sGen=sVoc sGen=pAcc sVoc=pNom sVoc=pAcc pNom=pVoc pAcc=pVoc;
- 10 < 15 > sVoc=sLoc ==> pNom=pVoc;
- 11 < 12 > sVoc=pNom ==> pNom=pVoc;
- 12 < 13 > sVoc=pAcc ==> pNom=pVoc;
- 13 < 5 > sVoc=pIns ==> sGen=pAcc sDat=sLoc pNom=pVoc masc;
- 14 < 3 > sLoc=sIns ==> sNom=sVoc sGen=sAcc sVoc=pNom sVoc=pAcc pNom=pVoc pAcc=pVoc;
- 15 < 10 > sLoc=pNom ==> pNom=pVoc;
- 16 < 7 > sLoc=pIns ==> sGen=pAcc pNom=pVoc;
- 17 < 5 > sIns=pGen ==> sGen=pAcc pNom=pVoc pAcc=pVoc fem;
- 18 < 4 > sIns=pDat ==> sNom=sVoc pNom=pVoc;
- 19 < 7 > sVoc=pAcc sLoc=pNom pNom=pVoc ==> sNom=sAcc;
- 20 < 7 > sVoc=pNom sLoc=pNom pNom=pVoc ==> sVoc=sLoc;
- 21 < 10 > sVoc=pNom sVoc=pAcc pNom=pVoc ==> pAcc=pVoc;
- 22 < 7 > sVoc=sLoc sLoc=pNom pNom=pVoc ==> sVoc=pNom;
- 23 < 6 > sVoc=sLoc sVoc=pAcc pNom=pVoc ==> sNom=sAcc sVoc=pNom sLoc=pNom pAcc=pVoc;
- 24 < 7 > sVoc=sLoc sVoc=pNom pNom=pVoc ==> sLoc=pNom;
- 25 < 3 > sGen=pAcc sLoc=pNom sLoc=pIns pNom=pVoc ==> sGen=sAcc pVoc=pIns masc;
- 26 < 10 > sGen=pAcc sVoc=pAcc pNom=pVoc ==> sGen=sVoc;
- 27 < 2 > sGen=pAcc sVoc=pNom sLoc=pIns pNom=pVoc ==> sGen=sAcc sDat=sLoc sDat=pIns;
- 28 < 3 > sGen=pAcc sVoc=sLoc sLoc=pIns pNom=pVoc ==> sDat=sLoc sDat=pIns sVoc=pIns masc;
- 29 < 6 > sGen=pAcc sDat=pIns sLoc=pIns pNom=pVoc ==> sDat=sLoc;
- 30 < 3 > sGen=pAcc sDat=pIns sLoc=pNom pNom=pVoc ==> sGen=sAcc;
- 31 < 3 > sGen=pAcc sDat=pIns

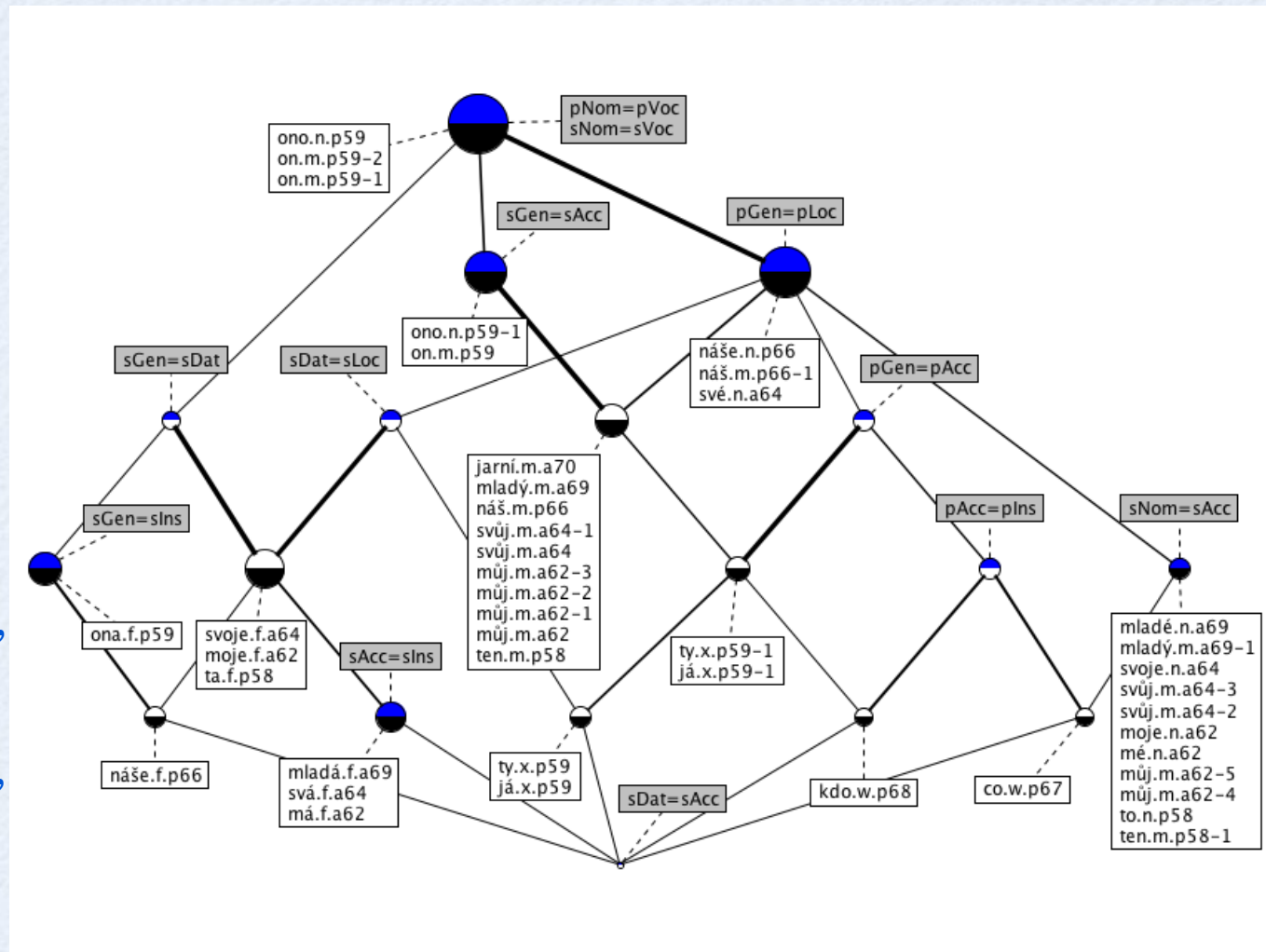
更に $sX=pY$, $pX=sY$, $s[^{(Nom)}]=sVoc$, $p[^{(Nom)}]=pVoc$

を除いた13属性

- 91個中の13個 (= $1/7 = 0.142$)
 - $sNom=sAcc$
 - $sNom=sVoc$
 - $sGen=sDat$
 - $sGen=sAcc$
 - $sGen=sVoc$
 - $sGen=sIns$
 - $sDat=sAcc$,
 - $sDat=sLoc$
 - $sAcc=sIns$
 - $sAcc=pAcc$
 - $sVoc=sLoc$
 - $sLoc=sIns$
 - $pNom=pVoc$
 - $pGen=pAcc$,
 - $pGen=pLoc$
 - $pAcc=pVoc$
 - $pAcc=pIns$
 - $pVoc=pIns$

代名詞と形容詞のFCA 1/2

- 論文の Figure 2
- 空所数0で過剰分類なし
- 過少分類なし
- 使用属性は次の12個
 - pNom=pVoc, sNom=sVoc,
 - sGen=sAcc, pGen=pLoc,
 - sGen=sDat, sDat=sLoc,
 - pGen=pAcc, sGen=sIns,
 - pAcc=pIns, sNom=sAcc,
 - sAcc=sIns, sDat=sAcc
- sLoc=sIns のみ除去



F=G の Optimality Theory 流の解釈

- F=G は anti-syncretism 条件 (*F=G) の違反なので
 - それがHasse 図の上にある程, 違反は軽度
 - それがHasse 図の下にある程, 違反は重度
- 具体的に言うと
 - *sDat=sAcc は中和不可能
 - *pNom=pVoc, *sNom=sVoc は (同じ程度に) 中和可能

代名詞と形容詞のFCA 2/2

- 参考 (論文にないHasse図)

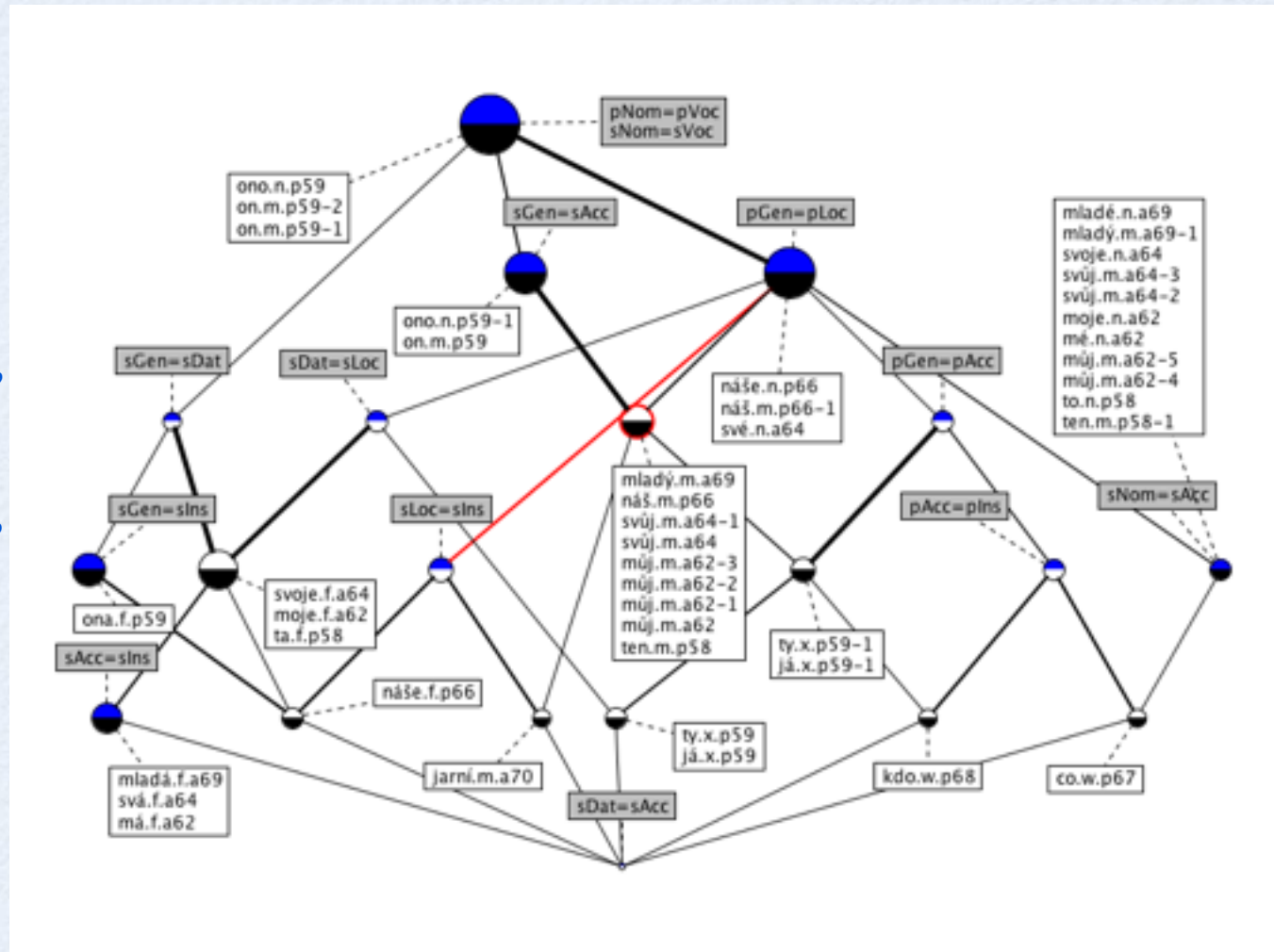
- 使用属性は次13個全部

● pNom=pVoc,
 sNom=sVoc,
 sGen=sAcc, pGen=pLoc,
 sGen=sDat, sDat=sLoc,
 pGen=pAcc, sGen=sIns,
 pAcc=pIns, sNom=sAcc,
 sAcc=sIns, sDat=sAcc,
sLoc=sIns

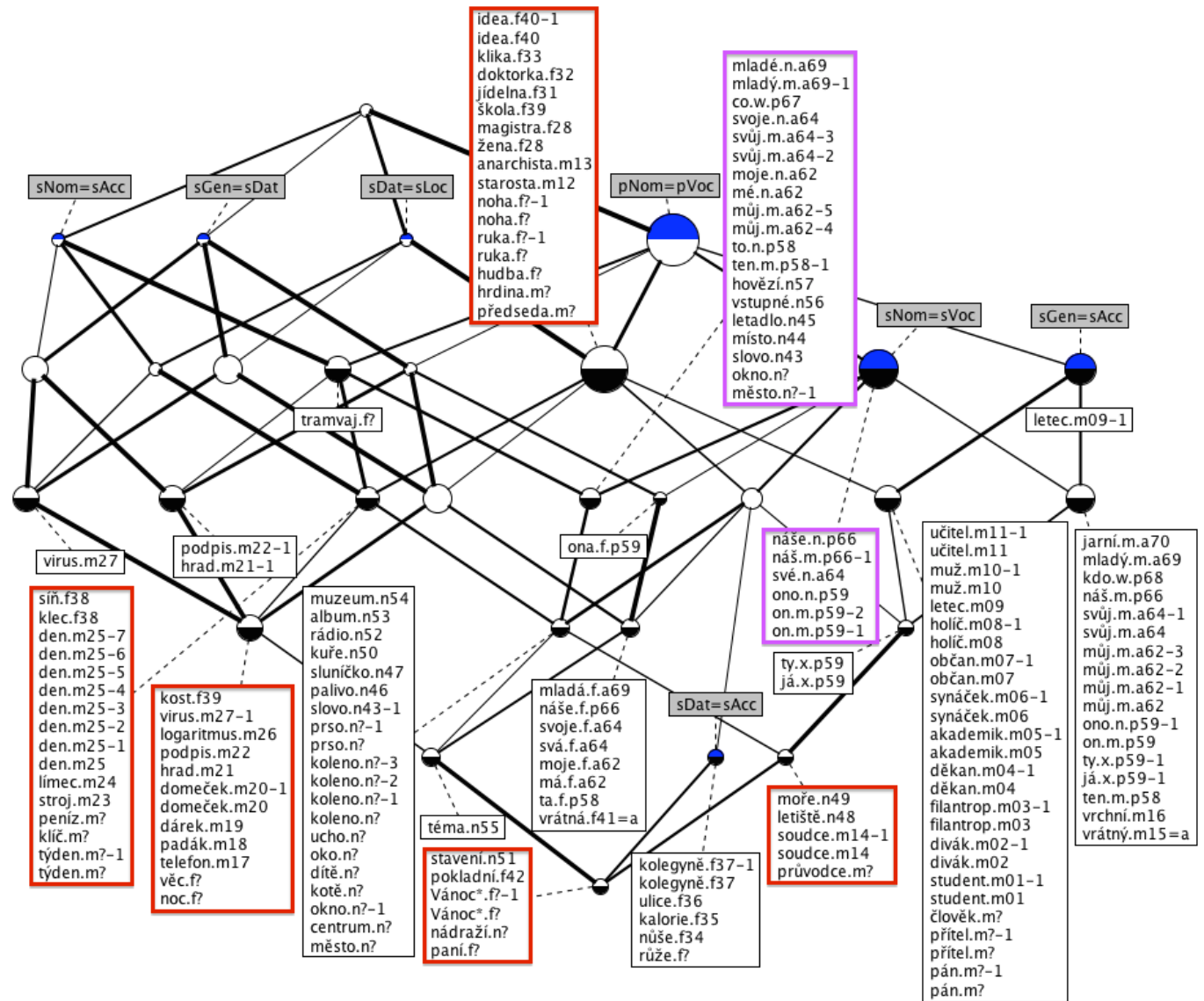
- 過剰分類. 過少分類なし

- 局所的破綻あり

- jarní.m.a70 が異例

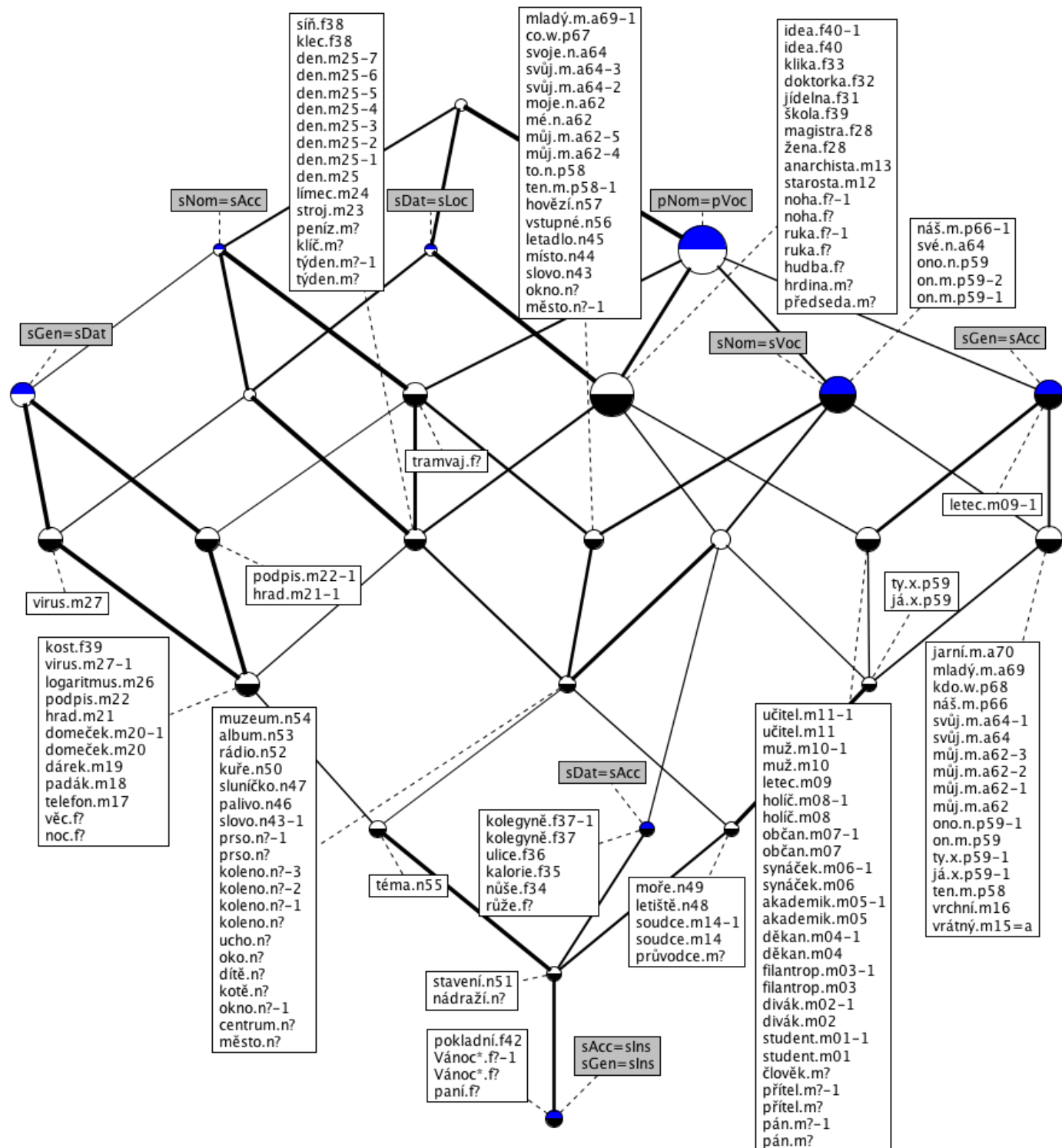


- 論文の Figure 3
- 全対象 (=代名詞, 形容詞を含む)の分離を(破綻を避けつつ)最大化した FCA
- 色のついた箇所が過少分類 = gender overload あり
- 使用属性は次の7個
 - sNom=sAcc, sGen=sDat, sDat=sLoc, pNom=pVoc, sNom=sVoc, sGen=sAcc, sDat=sAcc

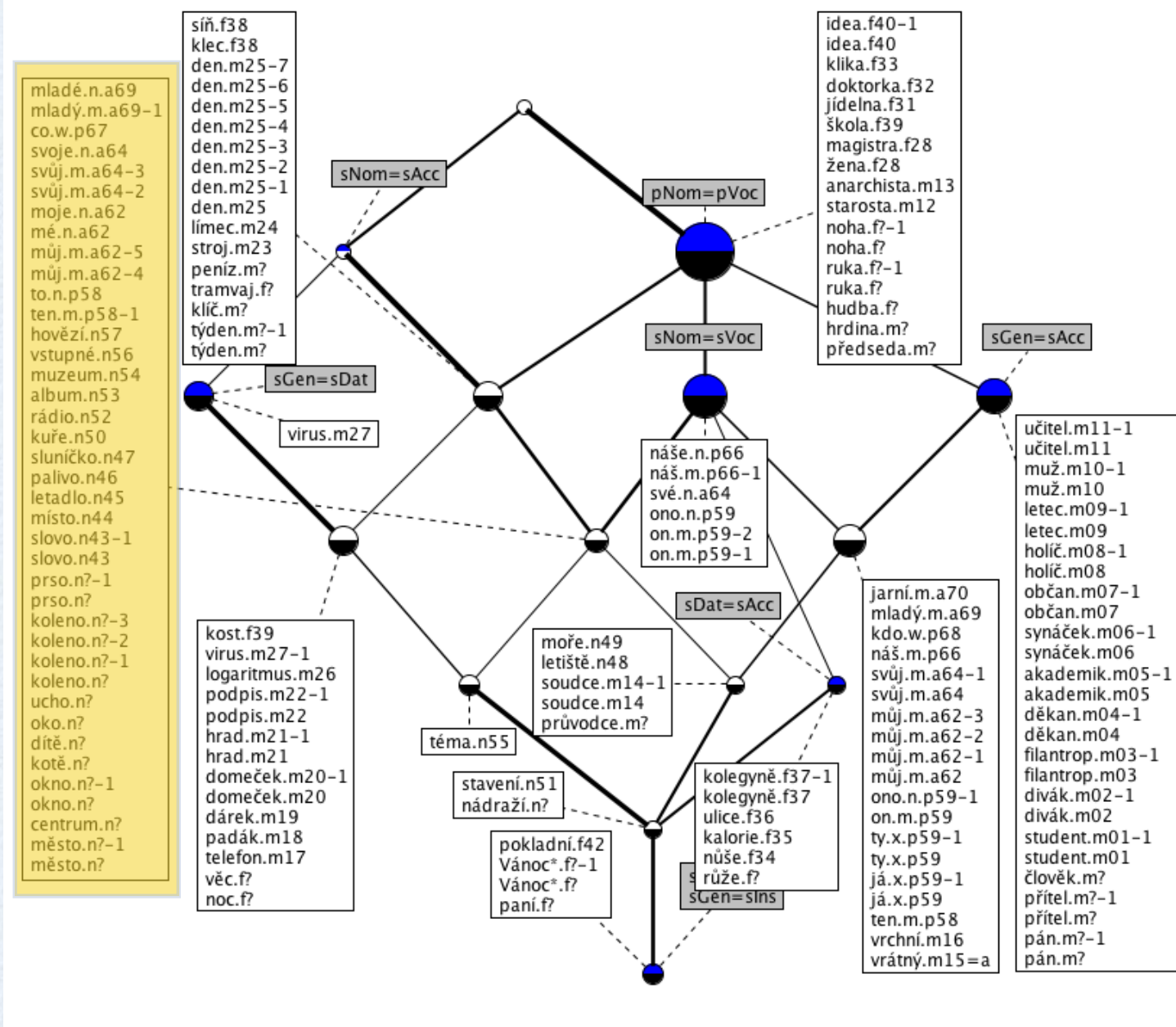


(c) Kow Kuroda: How are inflectional paradigms represented? at NLP22, 2016

- 論文の Figure 4
- 代名詞と女性形容詞を除いた対象のFCA
- 空所数は2
- gender overload を除いて過少分類なし
- 使用属性は次の9個
 - sNom=sAcc, sDat=sLoc, pNom=pVoc, sGen=sDat, sNom=sVoc, sGen=sAcc, sDat=sAcc, sAcc=sIns, sGen=sIns

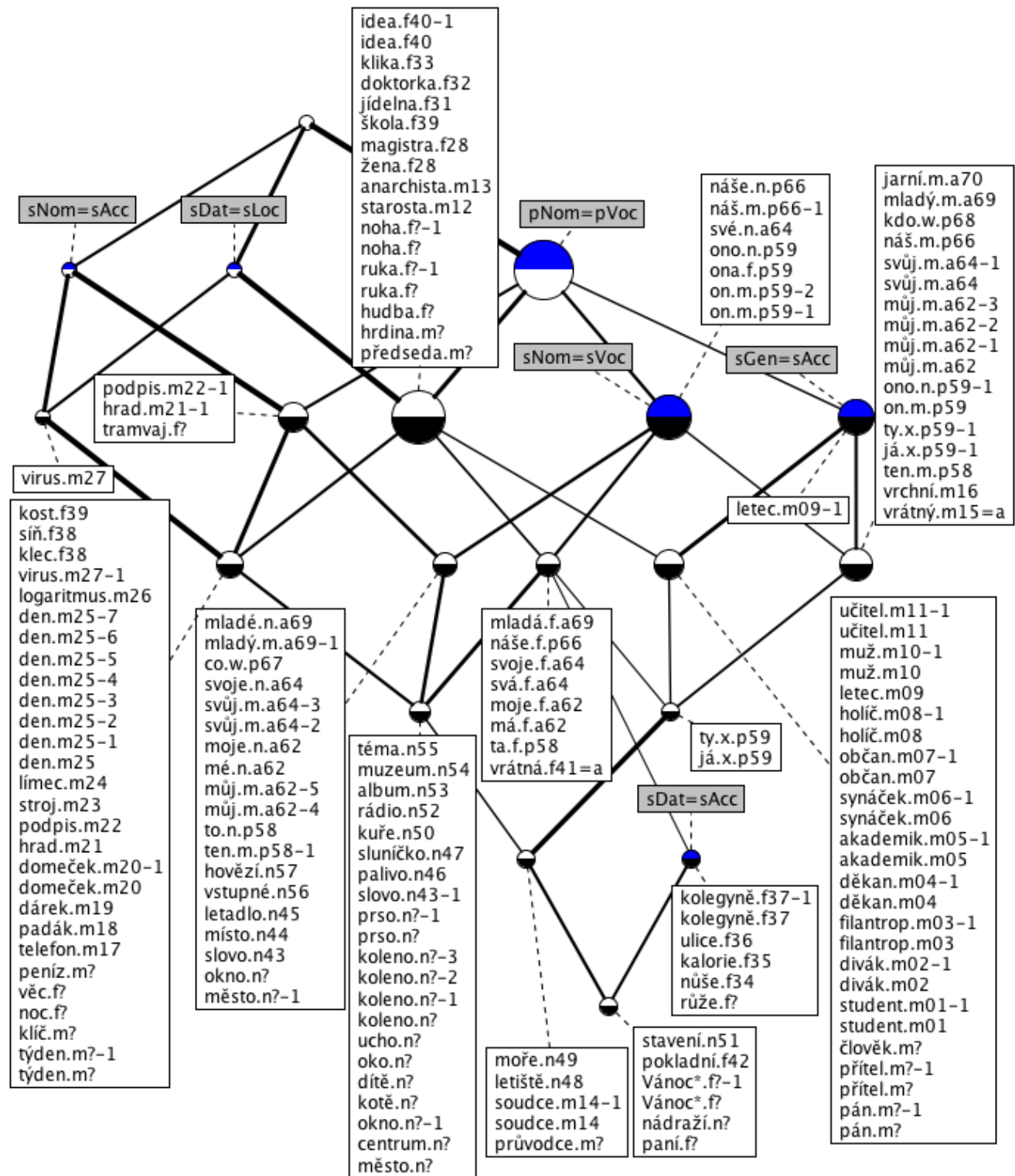


- 論文の Figure 5
- 女性代名詞, 形容詞のみを除いて
- 過剰分類を最小化
- 過剰分類は Masc. inanim と Neuter の同一視
- 黄色の集合
- 使用属性は次の8個
 - sNom=sAcc,
 - pNom=pVoc,
 - sGen=sDat, sNom=sVoc,
 - sGen=sAcc, sDat=sAcc,
 - sGen=sIns, sAcc=sIns



(c) Kow Kuroda: How are inflectional paradigms represented? at NLP22, 2016

- 論文にないHasse図
- 全 (=代名詞と女性形容詞を除かない) 対象で空所数 0 実現したFCA
- 使用属性は次の6つ
 - sNom=sAcc,
 - sDat=sLoc,
 - pNom=pVoc,
 - sNom=sVoc,
 - sGen=sAcc, sDat=sAcc
- 性別の overload はある



簡単な考察 1/2

- 代名詞は(挙動から見る限り)名詞と言うより形容詞
- 曲用クラスは(部分的には)意味クラスと相関
 - 本来は音韻的なものだと思われませんが

簡単な考察 2/2

- *F=G の anti-syncretism 条件には
 - 優先順位がある
 - Hasse図の下にある程, 違反が重度
 - Hasse図の上にある属性は目立つけれど, 体系を説明する力は弱い
 - 位置によって異なる説明上の役割を持つ
 - Hasse図の上にある属性は類似性の判断材料になる

まとめ

- Pairwise identity network (PIN) というデータ圧縮法を使って、チェコ語の名詞と形容詞の曲用クラスの (半自動) 検出に成功
 - Formal Concept Analysis を使って
 - FCA の実装は Concept Explorer 1.3
- 曲用クラスの階層化に成功
 - *F=G の制約の順位づけ (=重みづけ) に成功
 - 名詞, 代名詞, 形容詞の曲用の間の質的違いの検出
 - 名詞の内部での, 男性, 中性, 女性の中の曲用の質的違いの検出

今後の課題

- *F=G の優先順位の説明
- Static な側面
 - Paradigm Economy Principle の先行研究 [Ackerman et al. 2009, 2013; Blenvins 2013; Carstairs-McCarthy 1992, 1994] との対照を含む
- Dynamic な面
 - 言語進化のシミュレーションの研究成果 [Kirby and Hurford 2002] や Generative Social Science の研究成果 [Epstein 2006] との対照

References

- Ackerman, F., J. Blevins, & R. Malouf. 2009. Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter. In J. Blevins & J. Blevins (eds.), *Analogy in Grammar: Form and Acquisition*. Oxford: Oxford University Press.
- Ackerman, F., & R. Malouf. 2013. Morphological organization: The Low Conditional Entropy Conjecture. *Language*, 89, 429–464.
- Blevins, J. 2004. Inflection classes and economy. In G. Müller, L. Gunkel and G. Zifonun (eds.), *Explorations in Nominal Inflection*. Berlin: Mouton de Gruyter.
- Blevins, J. 2013. The information-theoretic turn. *Psihologija* 46: 355–375.
- Carstairs-McCarthy, A. 1992. *Current Morphology*, Routledge.
- Carstairs-McCarthy, A. 1994. Inflection classes, gender, and the Principle of Contrast. *Language* 70: 737–88.
- Epstein, J. M. 2006. *Generative Social Science: Studies in Agent-based Computational Modeling*. Princeton: Princeton University Press.
- Ganter, B. and R. Wille. 1999. *Formal Concept Analysis: Mathematical Foundations*, Springer Verlag.
- Ganter, B., G. Stumme, & R. Wille, eds. 2005. *Formal Concept Analysis: Foundations and Applications*, Springer Verlag.
- Kirby, S., & J. Hurford. 2002. “The emergence of linguistic structure: An overview of the iterated learning model.” In A. Cangelosi & D. Parisi (eds.) *Simulating the Evolution of Language*. London: Springer Verlag.
- McCarthy, J. 2008. *Doing Optimality Theory: Applying Theory to Data*, Wiley-Blackwell.
- Prince, A. and P. Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*, John Wiley & Sons.
- Wurzel, W.~Ulrike. 1987. System-dependent morphological naturalness in inflection. In Dressler, W.~U., Mayerthaler, W., Panagle, O., & Wurzel, W.~U. (eds.), *Leitmotifs in Natural Morphology*, pp .59–96. Berlin: Mouton de Gruyter.
- 鈴木 治 & 室伏 俊明. 形式概念分析: 入門・支援ソフト・応用. 知能と情報 19 (2): 103--142.

ご静聴に感謝

付録

全属性を使った全対象のFCA

