

**SwedishFrameNet (SweFN)**

**Språkbanken  
University of Gothenburg**

**December 2009**

**Arbetsrapport**

**Färdigställd av:  
Maria Toporowska Gronostaj  
Dana Dannélls**

# 1. Beskrivning av databasen

## 1.1. Fälttyper

Textfält i databasen: FR, FR\_Example, FR\_SemType, FR\_Domain, FE\_Core\_List, CFE\_ST\_Example, FE\_NonCore\_list, FE\_Cmp\_List, FE\_Cmp\_Example, LUs\_SALDO, LUs\_new, Notes, Created\_by.

## 1.2. Fälltsinnehall

Syntax gällande fältnamnen: db//html

### **FR (Frame) // ram**

Rubriker på ramar är unika i hela FrameNet. Det finns fyra fall för hur vi väljer ramarnas rubriker: (1) om ändringen sker enbart bland de lexikaliska enheterna behåller vi samma rubrik som i det EngFN, (2) om namnet på ramen i det EngFN är identiskt med ett ramelement (FE), läggs ordet *frame* till rubriken (t.ex. i det EngFN förekommer *Artifact* ramen som innehåller ett element med namnet *Artifact*, därför väljer vi att döpa om ramen i det SweFN till *Artifact\_frame*), (3) om vissa modifieringar görs bland de obligatoriska elementen som finns in det EngFN, t.ex. borttagning av ett ramelement, läggs ordet *reviderat* till rubriken, (4) vid skapandet av en ny ram väljer vi en rubrik som inte finns i det EngFN.

Varje ram illustreras med minst ett exempel som anges i fältet FR\_Example där det bärande predikatet/ordet (target LU) markeras explicit med LU och sätts inom hakparenteser '[]'. Även andra ramelementen markeras beträffande deras semantiska roller. T.ex. markeras ramelementen i ramen Experience\_bodily\_harm i satsen *han har skadat* på följande sätt [E Han] [LU har skadat] [B handen]. Om det rambärande ordet (target LU) är identiskt med något element markeras detta elementet på samma sätt som LU, dvs. namnet på ramelementet utelämnas (jfr ramen Medical\_professionals).

När det behövs att införa flera markeringar inom ett komplext ramelement (tex *sin arm*) så används följande konvention: [Body\_part [Experiencer sin] arm] (jfr ramen Experience\_bodily\_harm).

Vi inför en speciell markering för sammanflätning/conflation och stödverb/support verb (se The Book 52-53 för substantiv).

För sammanflätningen används P/A, tex. [LU cure] [P/A epileptic].

För stödverb används SUPP tex [SUPP ta] [LU medicin]

För copula används COP tex [COP vara] [SUPP på] [LU bättringsvägen].

### **FR\_Example (FR\_EX) // exempel**

Fältet kan innehålla ett eller flera exempel. Varje exempel separeras med en punkt eller annan interpunktionsmarkör om det är en sats, ingen markeringer för fraser. Dubbla semikolontecknen ska alltid stå sist i detta fält såväl som mellan varje exempel. Valet av exempel bör vara representativt för de i textkorporus förekommande semantiska och syntaktiska varianterna.

**FR\_ST (Semantic Type) // semantisk typ**

Detta fält är avsett att fånga en eventuell indikation av den semantiska typen angiven i EngFN och semantiska typer postulerade för SweFN med utgångspunkt i SIMPLE-ontologin. Notationen 0//0 används i odefinierade eller oklara fall.

**FR\_Domain (FR\_Dom) // domän**

Fältet fångar domänprofilen på LU\_listan, om orden förväntas vara frekventa i domänspecifika texter. Domänen bestäms utifrån ramen och dess CFEs (Core Frame Elements). Varje ram associeras med minst en profil. De tre profiler som vi använder för nuvarande är: Med (Medical), Art (Artwork), Gen (General). Om de lexikala enheter som anges i ramen kan förankras i lika stor utsträckning i både en specifik domän och ett allmänt språk markeras båda varianter, t.ex. i ramen *Death* innehåller FR\_Dom fältet Med/Gen eftersom de lexem som är representativa för ramen *Death* kan också förkomma i andra domäner, t.ex. i Art-relaterade texter.

**FE\_Core\_List // kärnelement**

Fältet innehåller en lista med obligatoriska kärnelement, som för det mesta är identiska med elementen i det EngFN, och deras förkortningar som ges inom parentestecken. Än så länge använder vi samma element och elementnamn som i det EngFN. Elementen separeras med ett komma. Andra kärnelement kan föreslås för nya SweFN ramar om ett sådant behov föreligger.

**CFE\_ST\_Example (CFE\_ST\_EX) //exempel**

(CFE, core frame element, kärnelement.) Dubbla semikolon används mellan alla exempel som illustrerar ett specifik CFE och varje CFE- rad avslutas med dubbla semikolon.

**FE\_NonCore\_list //periferielement**

Fältet innehåller en lista med icke-obligatoriska element och deras förkortningar som ges inom parentestecken. Elementen separeras med ett komma.

**FE\_Cmp\_List // sms**

I detta fält specificeras möjliga typer av sammansättningar som kombinerar CFE och NonCFE. CFE ska förekomma i en slutposition i sammansättningen. Varje kombination separeras med ett komma.

**FE\_Cmp\_Example // sms-exempel**

Varje kombination som förekommer på FE\_Cmp\_listan måste förekomma i det här fältet med ett eller flera exempel, dvs. relevanta sammansättningar. Varje sammansättning separeras med ett komma, sammansättningstyper separeras med dubbla semikolon. Dubbla semikolontecknen ska alltid stå sist i detta fält.

**LUs\_SALDO // saldo**

Fältet innehåller listor på de lexikala enheterna som tillhör ramen. Varje ordklass separeras med ett semikolon. Ordklass taggarna är identiska med de taggarna i SALDO, dvs. verb (vb), substantiv (nn) och adjektiv (av). Semikolontecknet används för att dels skilja åt LUs listor för de enskilda ordklasserna, dels för att markera listans slut.

**LUs\_new // saldo (nya)**

Här listas alla de lexikala enheterna som tillhör ramen men som saknas i SALDO. Fältets innehåll har samma syntax som gäller i LUs\_SALDO fältet.  
Dubbla semikolontecken ska alltid stå sist i detta fält efter varje ordklass kategori.

### **Notes // kommentar**

Fältet står för fria kommentarer. Man bör notera ramens status, särskilt för nya ramar anges ramens förhållande till en motsvarande engelsk ram.

## **1.3. Språk**

Namnen på semantiska ramar och element är på engelska, de flesta bär samma namn som i det EngFN. Exempelen och de lexikala enheterna är på svenska.

## **2. Semantiska ramar**

Databasen innehåller XX svenska ramar (X Med, X Art, X Gen), vi har X förslag på nya lexikala enheter som saknas i SALDO (X substantiv, X verb och X adjektiv)

### **2.1. Nya ramar**

Nya ramar redan införda i SweFN:

- Medical\_disorders (Jfr Medical\_conditions i EngFN)
- Health\_status (Jfr Medical\_conditions i EngFN)
- Falling\_Ill
- Timespan\_preliminary

### **2.2. Planerade ramar:**

Maria:

En ny ram som vore ett komplement till Observable\_bodyparts and People\_by\_disease Svenska rammotsvarigheter till Fall\_asleep, Feeling, Intoxicants, Intoxication, Inspecting, Scrutiny, Scrutinizing\_for, Verification, Removing, State\_of\_entity, Dead\_or\_alive, Dying, Separating, Part\_whole

Dana:

Arbeta vidare med Timespan ramen, Categorization, Title, People\_by\_vocation, Means, Locale\_by\_use, Document, Substance

## **3. Kopplingar mellan SweFN och andra lexikaliska resurser**

- SALDO
- SIMPLE ontologi
- IDS/LWT (Intercontinental dictionary series/Loan word typology)

## **4. Teoretiska frågeställningar och tankegångar**

### **4.1 Frågeställningar**

- Hantering av sammansättningar som består av olika FE
- Beskrivning av ramar och FE i termer av semantiska typer enl. FN//SIMPLE.

- Domän information
- Nya ramar
- Genomgång av ramar utifrån ett onomasiologiskt perspektiv (tex Medicin, Time-relaterade ramar) för att fylla i eventuella luckor i det EngFN,
- Undersökning av sammanflätade element ("conflation of core elements; The Book s. 25; t.ex. *cure the epileptic* conflation of Patient and Affliction)
- Propagationen av informationen om semantiska typer samt mappningen av OWL och FN (se The book. s.127)
- Möjlig sammanfogning av ramar utifrån (1) en ontologisk klassifikation, betydelse nivå, (2) lexikalisk, lexemnivå.

## 4.2. Tankegångar

### APPROACH

Den övergripande approachen till rambeskrivningen bygger på en kombination av ett onomasiologiskt och semantiskt perspektiv. Genom att arbeta utifrån domäner och deras frågeställningar kommer vi att få en mer uttömmande och kunskapsrelaterad beskrivning av relevanta relationer och begrepp för varje domän. Metoden möjliggör att fylla i luckor i den befintliga förteckning av engelska ramar, t.ex. *Falling\_ill* postuleras som ett komplement till *Recovery*, *People\_by\_diseases* som ett komplement till *Medical\_professionals*, *People\_by\_age* or *People\_by\_religion*.

En systematisk jämförelse och genomgång av typerna av semantiska element planeras för att kunna fånga möjliga generaliseringar och därigenom reducera antalet ramelement.

Taggningsystemet bygger på den EngFN., dvs. i substantiviska ramar taggar vi enbart nominala fraser som innehåller det substantiviska LU. I ramar där huvudordet är ett verb taggas alla argument som hör till verbet.

### CONFLATION

PATIENT/AFFLICTION: patient vars namn syftar på en typ av sjukdom, t.ex. reumatiker implicerar sjukdomen reumatism som sorteras under ramen *Health\_status*. Botandet av person innefattar oftast botandet av sjukdomen. Sådan information bör man kunna generera automatiskt ur ordböcker, dels från definitioner, dels med hjälp av ordbildningsregler (tex. NEO: reumatiker person som lider av reumatism).

REPRESENTED/REPRESENTATION: skapandet av tavlans motiv inbegriper/implicerar skapandet av en tavla.

Man bör fundera på: (1) till vilken grad beror sammanflätning på verbet (bota är OK, men ej behandla eftersom man kan t.ex. behandla reumatikern för andra åkommor). (2) om med hjälp av logiskt motiverade ramar (OWL) kan vi hitta potentiella fall av sammanflätning/conflation.

## 5. Teknisk specifikation

### 5.1. Den nuvarande versionen (v0.1)

- en ODB (Open Office DB) databas i html format finns i subversion

- ODB tillåter att formulera sql-frågor , dock inte så avancerade
- tillgång till den EngFN-databasen via Berkeley webbtjänsten

## 5.2. Framtidsutveckling

- inspektion av data utifrån vissa semantiska fält el. fältkombinationer
- tillgång till det EngFN-databasen (Dana har redan skrivit på licensavtalet, vi kommer snart att ha tillgång till hela databasen lokalt)

## Appendix A. Förkortningar

**CFE** Core Frame Element

**Cmp** Compound

**FR** Frame Name

**FE** Frame Element

**EngFN** English FrameNet

**ST** Semantic Type

## Appendix B. IDS/LWT concept

- 1 The physical world |
- 2 Kinship |
- 3 Animals |
- 4 The body |
- 5 Food and drink |
- 6 Clothing and grooming |
- 7 The house |
- 8 Agriculture and vegetation |
- 9 Basic actions and technology |
- 10 Motion |
- 11 Possession |
- 12 Spatial relations |
- 13 Quantity |
- 14 Time |
- 15 Sense perception |
- 16 Emotions and values |
- 17 Cognition |
- 18 Speech and language |
- 19 Social and political relations |
- 20 Warfare and hunting |
- 21 Law |
- 22 Religion and belief |
- 23 Modern world |
- 24 Miscellaneous function words