

# **The principle of compositionality and entity-relationship modelling: faceted classification in a broader context**

**Dagobert Soergel**

Department of Library and Information Studies  
Graduate School of Education  
University at Buffalo

UDC 2017

See the full paper for detail and references

# Intended take-aways

- 1 Understand compositionality as a pervasive principle in human thought and expression since very early on.**
- 2 Recognize application of compositionality in many KOS, faceted classification being just one.**
- 3 Understand that entity-relationship modeling, frames/objects in knowledge representation for AI and programming, and facets are but different expressions of the same underlying idea.**
- 4 Understand how entity-relationship modeling can be used to analyze and clarify relationships in UDC.**

# Outline

## 1. Introduction

## 2. The principle of compositionality at work in many contexts

2.1 Meaning from components

2.2 Systems using composition with more structure.  
Entity-relationship modeling, frames, and facets

## 3. Analyzing the structure of the UDC through entity-relationship (E-R) modelling

3.1 E-R model capturing the structure of some common auxiliaries

3.2 A matter of time. Use of an entity type across UDC

## 4. Conclusions

# Introduction

Compositionality and entity-relationship modelling are intertwined structural principles that underlie

- thought,
- language, and
- classification / knowledge representation / data modelling.

Drawing on examples from many contexts, this paper illustrates common principles for representing and understanding

- reality,
- imagination,
- conceptualization

in

- thought,
- natural language,
- systems designed for organizing and applying knowledge –
  - classification for organizing documents,
  - knowledge representation in AI and
  - data modelling for databases.

# Introduction

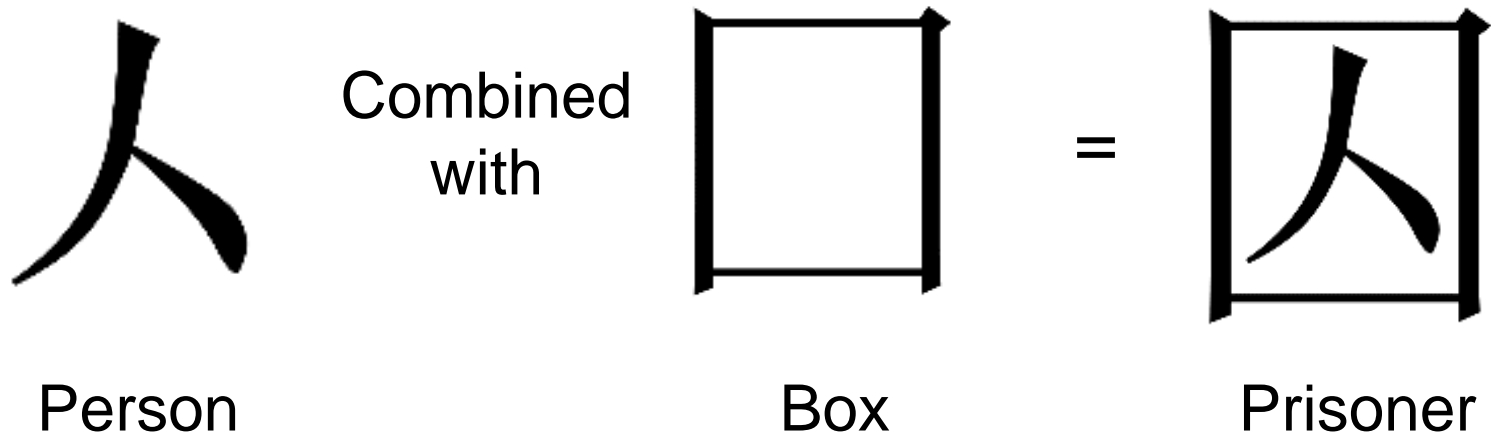
## The principle of compositionality

"The meaning of a complex expression is determined by its structure and the meanings of its constituents".

## **2. The principle of compositionality at work in many contexts**

# 2.1 Meaning from components

**Figure 1. Semantic composition in Chinese characters (3000 BC)**



# 2.1 Meaning from components

**Figure 2.** *Semantic composition in sign language characters (Lidell, 2013)*





Think

with

Same-As




Agree

# 2.1 Meaning from components

**Table 1:** *Examples from Laffal's concept dictionary*

**Table 1: Examples from Laffal's concept dictionary**

Categories		Dictionary entries	
<b>agre</b>	(agree, approve, cooperate, promise, subscribe)	Address	- n plac numr path
<b>anml</b>	(animals other than birds, fish, and insects)	address2	- n talk
<b>join</b>	(bond, congregate, touch, union)	adhere	- v agre join
<b>numr</b>	(algebra, score, statistic)	admiral	- n lead flow vehc mili
<b>thma</b>	(ideas and thought, as in analyse, define, think, understand)	Bilingual	- j talk numr
		corollary	- j join thma
		dreamland	- n plac erth myth thma rest
		encode	- v sign writ thma covr



**Table 2. Word meanings (complex concepts)  
represented as combinations of "semantic markers"  
(elemental concepts) (from Katz, 1964)**

## **bachelor**

- (1) Physical Object, Living, Human, Adult, Male, Never-married;
- (2) Physical Object, Living, Human, Youth, Knight, Serving under the standard of another;
- (3) Physical Object, Living, Human, Having academic degree for completion of four years college;
- (4) Physical Object, Living, Animal, Male, Seal, Without a mate at breeding time

# UDC Synthetic Structure

**Tourist maps of Grafton County (USA, Maine) from 1970s**

**348.48(734.211.4)"197"(084.3)**

**348.48**

Tourism [*main subject*]

**(734.211.4)**

Grafton County [USA, Maine] [*place*]

**"197"**

1970s [*time*]

**(084.3)**

Maps [*form*]

From the UDC introduction.

# The Semantic Code.

Developed at Western Reserve University by Perry & Kent (1958);  
Examples are from a review by Vickery (1959) providing a good  
sense of the system.

**Earthquake = CATC.GWRD**

C-TC Disaster

A Categorical:

*Earthquake is a species of Disaster*

G-RD Ground

W Instrumental:

*Earthquake is produced by, acted on by, or  
acts on Ground*

# Semantic Code

The brittleness of cermet materials as influenced by their microstructure  
=

Three components:

brittleness = **PAPR010**] P\_PR property, 010 spec.brittle, analytic infix **A** isa

cermet materials = **CERM.2X.METL** C\_RM Ceramics, M\_TL Metals, **E** intrinsic, 2X  
two ceramics

microstructure of metal = **RANG.CRS.MYTL**

R\_NG Arrangement **A** isa; C\_RS Crystal **I** Is component of;  
M\_TL Metal **Y** Possessor of attribute

components combined, giving for each its role(s)

**KWV. KAP. PAPR. 010 | KOV. CERM. 2X. METL. | KAL. RANG.CRS.MYTL**

Role Indicators:

**KWV** Property given, **KAP** Property influenced; **KOV** Property given for; **KAL** Influencer

**Table 3: Faceted arrangement of the Semitic-Greek alphabet  
(Clarke, 1884)**

Facet 1: Soft vs hard	Facet 2: How the sound is produced				
		labial	guttural	dental	
Soft 1 <sup>st</sup> half of alphabet	α alpha	β beta	γ gamma	δ delta	ε epsilon
Hard 2 <sup>nd</sup> half of alphabet	ο omicron	π pi	ρ rho (guttural)	σ sigma, τ tau	υ upsilon



# Table 4: Entity-relationship model for data about alphabet characters

Entity types with their values	Relationship types
CharacterSound SoundProductionMechanism labial; guttural; dental	Sound <producedBy> SoundProductionMechanism Sound <hasArticulationIntensity> ArticulationIntensity
ArticulationIntensity soft; hard	$\beta$ <producedBy> <b>labial</b> $\beta$ <hasArticulationIntensity> <b>soft</b>  $\pi$ <producedBy> <b>labial</b> $\pi$ <hasArticulationIntensity> <b>hard</b>

# Table 5: Entity-relationship model for describing coins (a rudimentary model)

Entity types with sample values	Relationship types		
Coin (a class of coins) US Penny 1982 (USP1982)	Coin	<i>&lt;isDenominationOf&gt;</i>	Currency
Currency USD, GBP, EUR, JPY, BRL	USP1982		USD
Number	Coin	<i>&lt;hasFractionOfUnitValue&gt;</i>	Number
Time (Point in time or time period) 1982	USP1982		.01
Shape circle, hexagon, octagon	Coin	<i>&lt;producedInTime&gt;</i>	Time
Length (in mm)	USP1982		1982
Material gold, copper, nickel	Coin	<i>&lt;hasShape&gt;</i>	Shape
	USP1982		circle
	Coin	<i>&lt;hasDiameter&gt;</i>	Length
	USP1982		19 mm
	Coin	<i>&lt;MainBodyMadeOf&gt;</i>	Material
	USP1982		nickel
	Coin	<i>&lt;isCoatedWith&gt;</i>	Material
	USP1982		copper

**Table 6: Facet frame for describing coins**

<b>Facet</b>	<b>Value</b>
<i>Coin</i>	US Penny 1982
<i>Currency</i>	USD
<i>Value</i>	.01
<i>YearProduced</i>	1982
<i>Shape</i>	Circle
<i>Diameter</i>	19 mm
<i>MainBodyMaterial</i>	nickel
<i>CoatingMaterial</i>	copper

## **2.2 Systems using composition with more structure. Entity-relationship modelling, frames, and facets**

**Table 7:** *Case frame for buy/sell (much simplified)*

**Table 7: Case frame for buy/sell (much simplified)**

<b>Case role</b>	<b>Entity type</b>	<b>Sample value</b>
<u>SourceOfGoodOrService</u> / <u>ReceiverOfMoney</u> :	LegalEntity	Safeway
<u>ReceiverOfGoodOrService</u> / <u>SourceOfMoney</u> :	LegalEntity	Fred
GoodOrServiceExchanged:	GoodOrService	Can of soup #3574
PriceReceived / PricePaid:	MoneyAmount:	\$1
Note: Further slots could be added, e.g., for time and place of the event		

## **2.2 Systems using composition with more structure. Entity-relationship modelling, frames, and facets**

### ***Pustejovsky's Generative Lexicon***

## 2.2 Systems using composition with more structure. Entity-relationship modelling, frames, and facets

### Lexical Data Structure

- a. **Lexical typing structure**      Giving an explicit type for a word positioned within a type system for the language
- b. **Argument structure**      Specifying the number and nature of the arguments to a predicate
- c. **Event structure:**      Defining the event type of the expression and any subeventual structure it may have
- d. **Qualia structure**      A structural differentiation of the predicative force for a lexical item

***Pustejovsky's Generative Lexicon*** uses a more elaborate approach. The frame defining the meaning of a word is structured as follows (Pustejovsky, 2014).

**Table 8: Corpus-driven type system** (adapted from Pustejovsky, 2014)

<b>Entity</b>	<b>Eventuality</b>
. Physical Object	. State
. . Inanimate	. . Abstract State
. . Animate	. . . Privilege
. . Plant	. . . Psych
. . Location	. . . . Attitude
. Institution	. . . . Emotion
. Energy	. . . . Goal
. Astract Object	. . . Time Point
<b>Eventuality</b> (see right col.)	. . . Obligation
Part	. . . Responsibility
Property	. . . Power
Group	. . . Uncertainty
	. . . Concept
	. . . . Proposition
	. . . . Narrative
	. . . . Information
	. . . . Rule
	. Event



## Pustejovsky's Qualia Structure roles (adapted from Pustejovsky 2014).

a. <b>Formal, what x is</b>	The basic category of which distinguishes the meaning of a word within a larger domain; encodes taxonomic information about the lexical item; is-a relation;
b. <b>Constitutive, what x is made of</b>	The relation between an object and its material, constituent parts; part-of or made-of relation;
c. <b>Telic, function of x</b>	The purpose or function of the object, if there is one; used-for or functions-as relation;
d. <b>Agentive, how x came into being</b>	The factors involved in the object's origins or "coming into being"; created-by relation.

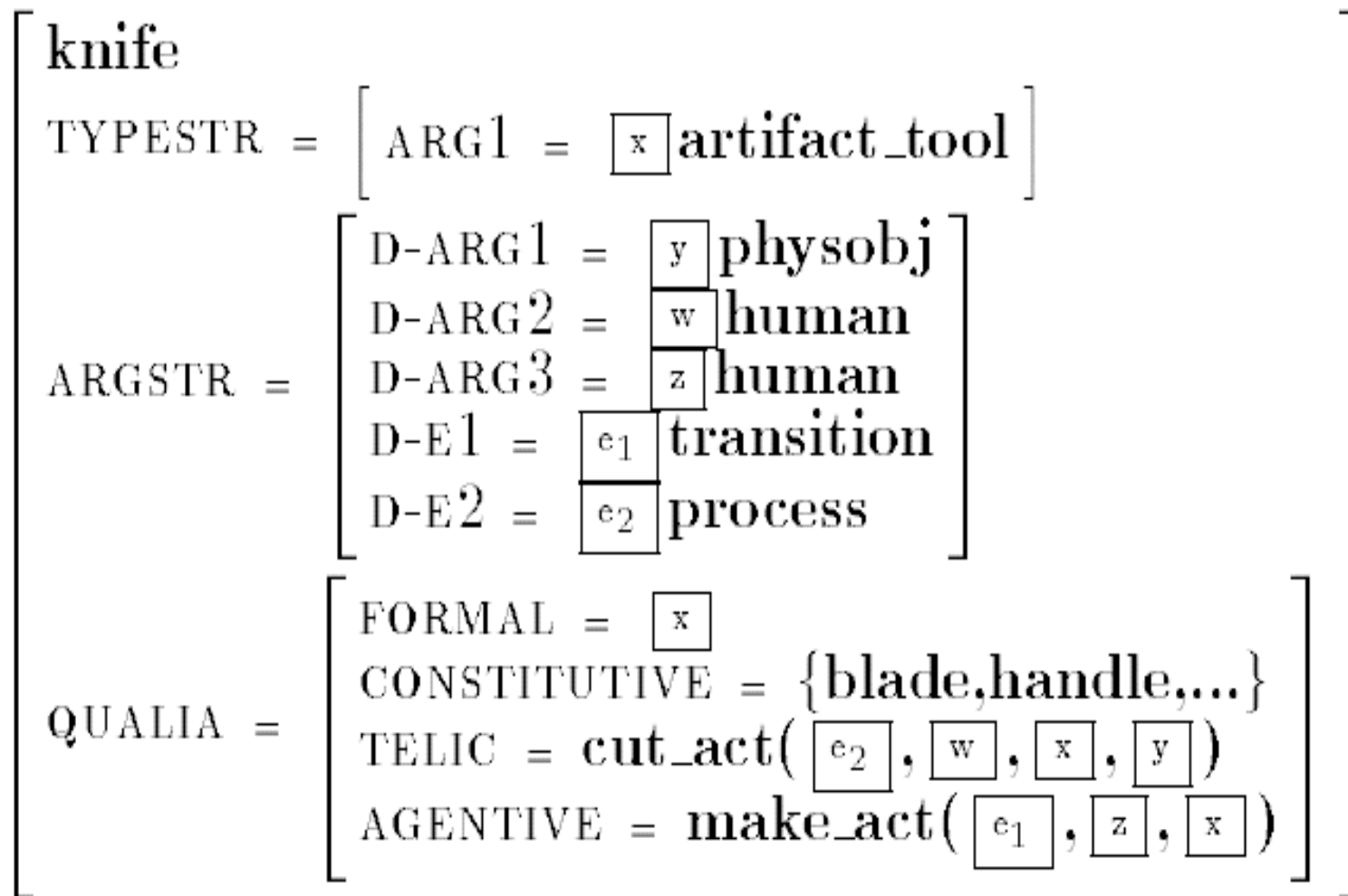
**Figure 3.** Generative Lexicon: example frames from Pustejovsky (1995) as reproduced in Weijnitz (2002).

$$\left[ \begin{array}{l}
 \textit{drive} \\
 \text{EVENTSTR} = \left[ \begin{array}{l}
 E_1 = \quad e_1:\text{process} \\
 E_2 = \quad e_2:\text{process} \\
 \text{RESTR} = \quad <^0_\alpha
 \end{array} \right] \\
 \text{ARGSTR} = \left[ \begin{array}{l}
 \text{ARG1} = \quad x:\text{human} \\
 \text{ARG2} = \quad y:\text{vehicle}
 \end{array} \right] \\
 \text{QUALIA} = \left[ \begin{array}{l}
 \text{FORMAL} = \quad \text{move}(e_2,y) \\
 \text{AGENTIVE} = \quad \text{drive\_act}(e_1,x,y)
 \end{array} \right]
 \end{array} \right]$$

**Figure 3.** Generative Lexicon: example frames from Pustejovsky (1995) as reproduced in Weijnitz (2002).  
continued

<i>Honda</i>	
ARGSTR =	[ ARG1 = x:car ]
QUALIA =	[ FORMAL = x TELIC = drive(e,y,x) AGENTIVE = create(e,Honda-co,x) ]

**Figure 3.** Generative Lexicon: example frame from Johnston & Busa (1999).



# **3. Analysing the structure of the UDC through entity relationship (E-R) modelling**

## 3.1 E-R Model Capturing the Structure of Some Common Auxiliaries

**Table 9:** *Examples of entity types and relationship types in UDC*

**Table 9:** *Examples of entity types and relationship types in UDC.*

Entity type	Relationship types
<b>UDC Table 1c - Common auxiliaries of language</b>	
The main table of languages . Note: Table 1c =02/03 seem out of place	
Language/ethnicity	Document <b>&lt;expressedIn&gt;</b> Language Document <b>&lt;dealsWith&gt;</b> Language WorkOfLiterature <b>&lt;expressedIn&gt;</b> Language Document <b>&lt;dealsWith&gt;</b> [WorkOfLiterature <b>&lt;expressedIn&gt;</b> Lang]  <b>UDC references</b> Table 1c =00, =1/=9; 811; 821

Entity type	Relationship types
<b>UDC Table 1c - Common auxiliaries of language (Table 9 Cont.)</b>	
The following can be used in combination with any language	
General features of languages . Table 1c =-...-2	
LivingStatus (Living, Dead, Revived)	Language <b>&lt;hasLivingStatus&gt;</b> LivingStatus Note: Depends on historical time <b>UDC references</b> Table 1c =...-23 =...-24, =...`08
LanguageStatusSpWr (Spoken, Spoken&Written WrittenOnly)	Language <b>&lt;hasLanguageStatusSpWr&gt;</b> LanguageStatusSpWr <b>UDC references</b> Table 1c =...-25 / =...-26



Entity type	Relationship types
<b>UDC Table 1c - Common auxiliaries of language (Table 9 Cont.)</b>	
The following can be used in combination with any language	
Sp. aux. subdiv. for origins and periods of lang. and phases of dvlpmnt Table 1c ...`01/`08	
Time (HistoricalPeriod)	Language <b>&lt;originatedIn&gt;</b> Time Language <b>&lt;usedIn&gt;</b> Time <b>UDC references</b> able 1c =...`01/`06
GroupOfPeople	Language <b>&lt;spokenBy&gt;</b> GroupOfPeople <b>UDC references</b> Table 1c =...`276
GeographicalScope (Local, Regional, Natl.)	Language <b>&lt;usedIn&gt;</b> GeographicalScope <b>UDC references</b> Table 1c =...`282

Entity type	Relationship types
<b>UDC Table 1d - Common Auxiliaries of form (Table 9 Cont.)</b>	
Special auxiliary subdivision for document form: physical features, production, use characteristics and supplementary matters of documents, Table 1d (0.02/.08)	
Size	Document <b>&lt;hasSize&gt;</b> Size <b>UDC references</b> Table 1d (0.021)
CountNumber	Document <b>&lt;hasNumberOfColours&gt;</b> CountNumber <b>UDC references</b> Table 1d (0.022.2/5)
DegreeOfTransparency	Document <b>&lt;hasDegreeOfTransparency&gt;</b> DegreeOfTransparency <b>UDC references</b> Table 1d (0.022.7/8)
Material	Document <b>&lt;madeOf&gt;</b> Material <b>UDC references</b> Table 1d (0.023)
VisuallImpressionType	Document <b>&lt;includes&gt;</b> Image Image <b>&lt;presentsVisuallImpressionType&gt;</b> VisuallImpressionType <b>UDC references</b> Table 1d (0.024)
CountNumberZeroMore	Document <b>&lt;includesIllustrations&gt;</b> CountNumberZeroMore <b>UDC references</b> Table 1d (0.025)

Entity type	Relationship types
<b>UDC Table 1d - Common Auxiliaries of form (Table 9 Cont.)</b>	
BindingType	Document <boundWith> BindingType <b>UDC references</b> Table 1d (0.026)
(EditionType) In turn faceted, needs more analysis	Document <has EditionType> EditionType <b>UDC references</b> Table 1d (0.027)
TypeOfDefect	Doc.<has TypeOfDefect> TypeOfDefect <b>UDC references</b> Table 1d (0.028)
(MethoOfProduction) Needs much analysis	Doc. <producedWith> MethodOfProduction <b>UDC references</b> Table 1d (0.03)
DocProductionStage	Doc <isInDocProductionStage> DocProductionStage <b>UDC references</b> Table 1d (0.04)
Group (see above, fn 3)	Document <suitableFor> Group <b>UDC references</b> Table 1d (0.05)

Entity type	Relationship types
<b>UDC Table 1d - Common Auxiliaries of form (Table 9 Cont.)</b>	
LevelOfExposition	Document <b>&lt;has LevelOfExposition&gt;</b> LevelOfExposition <b>UDC references</b> Table 1d (0.062/4)
Availability	Document <b>&lt;has Availability&gt;</b> Availability <b>UDC references</b> Table 1d (0.067/8))
Common auxiliaries of form Table 1d (01/9). Includes several facets.	
DocFormOfContent	Doc <b>&lt;has DocFormOfContent&gt;</b> DocFormOfContent <b>UDC references</b> Table 1d (01/3), (09)
Periodicity	Doc. <b>&lt;hasAppearancePattern&gt;</b> Periodicity <b>UDC references</b> Table 1d (05 in part)
Purpose	Document <b>&lt;servesPurpose&gt;</b> Purpose <b>UDC references</b> Table 1d (06/7)
Table 1d (0.07/8), (04) and (08) are mixed up and not easily analysed	

## 3.2 A Matter of Time.

# Use of an Entity Type Across UDC

### **Table 10:** *Use of Time in UDC*

Class and relationship	Comments
<b>Sections of UDC that give a breakdown of Time Periods. Time as a facet</b>	
<p>"..." Table 1g –  <b>Common auxiliaries of time</b></p> <p>Document <b>&lt;dealsWith&gt;</b>  Time<sup>1</sup></p>	<p>The main place for giving the values of the entity type Time. Used to specify the time component in many compound concepts built by combination, with implied relationships such as</p> <p>ThingOrIdea <b>&lt;wasProducedIn&gt;</b> Time</p> <p>ThingOrIdea <b>&lt;wasUsedIn&gt;</b> Time</p> <p>EventOrProcess <b>&lt;takingPlaceIn&gt;</b> Time</p>

Class and relationship	Comments
<b>Sections of UDC that give a breakdown of Time Periods. Time as a facet</b>	
<p>Table 1c =...`01/^06  <b>Special auxiliary subdivision for origins and periods of language</b> and phases of development</p>	<p>Dividing the Time Period that is generally considered "history" (as opposed to "prehistory") into four large Periods:          =...`01 Old period. Archaic period          =...`02 Classical period          =...`04 Middle period          =...`06 Modern period</p>
<p>7.03`01/^06          Phases of development in art</p>	<p>Follows Table 1c=...`01/^06 exactly except that 7.03`01/^06 is called Late period</p> <p>But also see below 7.031</p>

Class and relationship	Comments
<p align="center"><b>Sections of UDC that give a breakdown of Time Periods. Time as a facet Table 10:</b>  <i>Use of Time in UDC, Continued</i></p>	
<p>82`01/^06  Special auxiliary subdivision for periods and phases of developments of literature</p>	<p>Similar to, Table 1c =...`01/^06 but not exactly the same  82`01 Old or early periods. Earliest forms  82`04 Middle period. Mediaeval  82`06 Modern period</p>
<p>21 Prehistoric religions. Religions of early societies  212 Prehistoric religions  213 Religions of early societies</p>	<p>It appears that the religions beginning at 2.2 are considered to be newer. They are divided by other criteria.  212 further specified by a class from Table 1g "63"</p>
<p>003.315 Prehistoric writings. Embryo-writing. Protoliner writing</p>	
<p>551.7 Historical geology. Stratigraphy</p>	<p>Has time-related concepts but not actual values for the entity type Time</p>



Class and relationship	Comments
<p align="center"><b>Sections of UDC that give a breakdown of Time Periods. Time as a facet Table 10: Use of Time in UDC, Continued</b></p>	
<p>Document <b>&lt;dealsWith&gt;</b>Time is shorthand for  Document <b>&lt;dealsWith&gt;</b> Subject, Subject <b>&lt;hasComponent&gt;</b> Time</p>	
<p>551.583 Natural variations of climate. Climatic change</p> <p>551.583.2 Variations of climate in historical periods</p> <p>551.583.3 Prehistoric variations of climate.</p>	<p>Note the inconsistency of listing the newer period first</p>

**Table 10:** Use of Time in UDC, Continued

Class and relationship	Comments
<p align="center"><b>Sections of UDC that give a breakdown of Time Periods. Time as a facet Table 10:</b> <i>Use of Time in UDC, Continued</i></p>	
<p>Document <b>&lt;dealsWith&gt;</b>Time is shorthand for Document <b>&lt;dealsWith&gt;</b> Subject, Subject <b>&lt;hasComponent&gt;</b> Time</p>	
<p>7.01/.09 Special auxiliary subdivision for the arts 7.03 Artistic periods and phases. Schools, styles, influences 7.031 Prehistoric art. Primitive art (ancient and modern) 7.032 Art of the ancient cultures of the Old World. Styles of Antiquity</p>	<p>The broad classification as Prehistoric art does not require Table 1g, but one can use Table 1g to refine the age specification: ◇ 7.031.1 "632/633" Palaeolithic and Mesolithic art (Old and Middle Stone Age) ◇ 7.031.1 "634" Neolithic art (New Stone Age) ◇ 7.031.1 "636" Copper Age art  Note: 7.03 "636" also is Copper Age art, but would not arrange correctly</p>

Class and relationship	Comments
<p align="center"><b>Sections of UDC that give a breakdown of Time Periods. Time as a facet</b>  <b>Table 10: Use of Time in UDC, Continued</b></p>	
<p>Document <b>&lt;dealsWith&gt;</b>Time is shorthand for  Document <b>&lt;dealsWith&gt;</b> Subject, Subject &lt;hasComponent&gt; Time</p>	
<p>9            GEOGRAPHY.                BIOGRAPHY.                HISTORY</p> <p>902/908 Archaeology.  Prehistory. Cultural  remains. Area  studies</p> <p>903        Prehistory.  Prehistoric remains,  artefacts, antiquities</p> <p>904        Cultural remains of  historical times</p>	<p>Again, a coarse chronological subdivision in the main classes with the possibility of refining the subdivision by classes from Table 1g</p>

Class and relationship	Comments
<p align="center"><b>Sections of UDC that give a breakdown of Time Periods. Time as a facet</b>  <b>Table 10: <i>Use of Time in UDC, Continued</i></b></p>	
<p>Document <b>&lt;dealsWith&gt;</b> Time is shorthand for  Document <b>&lt;dealsWith&gt;</b> Subject, Subject &lt;hasComponent&gt; Time</p>	
<p>94 General history</p>	<p>Mostly combined with  Place (Table 1e) or  ethnic group (Table 1f),  then Time (Table 1g).   For each country specific historical subdivisions are  enumerated and given notations constructed using  the rules of Table 1g – very helpful and very clever.</p>

# 4. Conclusions

- 1 **Entity-relationship modelling is the basis of all methods of knowledge representation.**
- 2 **To design a faceted classification,**
  - first create an entity-relationship model
  - then represent this model as a facet frame.
- 3 **"Re-engineering" an E-R model** from a faceted classification can shed much light on the classification's structure.
- 4 Instructions for facet analysis can be reframed as instructions for constructing an E-R model and vice versa.

# 4. Conclusions 2

- 5 Assemble a **canon of entity types and relationship types (and therefore facets)** that occur across multiple domains to provide guidance

Be aware that other **entity types and relationship types (and therefore facets)** are very specific to a domain and must be discovered and defined for that domain (see Table 9).

**Following any rigid schema just stands in the way of adequately modelling the particularities and complexities of any domain.**

# 4. Conclusions 3

Combining the thinking about entity types, relationship types, and facets from both literatures enriches both.

Add insights from linguistics, and you have a really powerful conceptual apparatus

# Take-aways

- 1 Understand compositionality as a pervasive principle in human thought and expression since very early on.**
- 2 Recognize application of compositionality in many KOS, faceted classification being just one.**
- 3 Understand that entity-relationship modeling, frames/objects in knowledge representation for AI and programming, and facets are but different expressions of the same underlying idea.**
- 4 Understand how entity-relationship modeling can be used to analyze and clarify relationships in UDC.**



# Thank you

## Questions

**Dagobert Soergel**  
**ds@dsoergel.com**  
**www.dsoergel.com**