

Issues on UWs in the UNL System

I UNL Panel

COLING 2012, December 15th , Mumbai, India

Sameh Alansary

sameh.alansary@bibalex.org

Head of Linguistics Department, Alexandrina University.

Head of the Arabic UNL Language Center,

Bibliotheca Alexandrina,

Alexandria, Egypt.

Why asking these questions?

For pure linguistic purposes ?!

For language engineering purposes ?!

Answering these questions...

```
graph TD; A[Answering these questions...] --> B[Computational Linguistic point of view]; A --> C[Cognitive point of view]; A --> D[Different languages point of view];
```

Computational
Linguistic point of view

Cognitive point
of view

Different languages point
of view

Question 1:

How many UWs should be recognized in the sentence below?

"Charles Dickens is generally regarded as the most important English novelist of the Victorian period"

To answer this question:

- ✓ What is a UW?
- ✓ What is a concept?
- ✓ Open and closed lexical categories.
- ✓ Compositionality, predictability and functionality.

What is a UW ??

- UNL: UWs must correspond only to semantic discrete units conveyed by natural language open lexical categories (noun, verb, adjective and adverb).
- Any other semantic content (such as this conveyed by articles, prepositions, conjunctions etc.) should be represented as Attributes or Relations.
- If a given semantic value proves to be conveyed, in any language, by a closed class, it should not be represented as a UW, regardless of its realization in other languages (*not language-biased*).

What is a concept ??

Three most common understandings of the concept

- Concepts as mental representations: where concepts are mental symbols or representations.
- Concepts as abilities: where concepts are abilities peculiar to cognitive agents.
- Concepts as abstract objects: where objects are the constituents of propositions that mediate between thought, language, and referents.

Important Questions

Do all content words
(open class) represent
concepts or UWs ???

Are all function words
the same in different
languages???



Victorian period" should be represented as single UW or as two UWs?

Compositionality and Predictability

Compositional → Predictable

Example: Swimming pool

Non-Compositional → Unpredictable

Example: Cloud nine

In between → Semi-predictable

Example: Zebra crossing

Compounding in different languages

- Not all languages use compounding with the same frequency or flexibility.

Example:

English compound

French translation

steel knife



couteau en acier

cheese knife

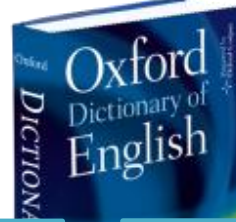


couteau `a fromage

Victorian



Period



Definition of **Victorian adjective**

- relating to the reign of Queen Victoria.

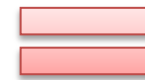
Definition of **period noun**

- a portion of time in the life of a nation, civilization, etc. characterized by the same prevalent features or conditions

Meaning of word 1



Meaning of word 2



Meaning of compound

Answer: “*Victorian period*” should be represented as two different UWs.

*The verb "to be" should be represented as a UW or as a relation?
(Consider "was", "has been" in the same context)*

➤ **What category of verbs does the verb "to be" belongs to ??**

Definition of be

- (usually **there is/are**) exist:

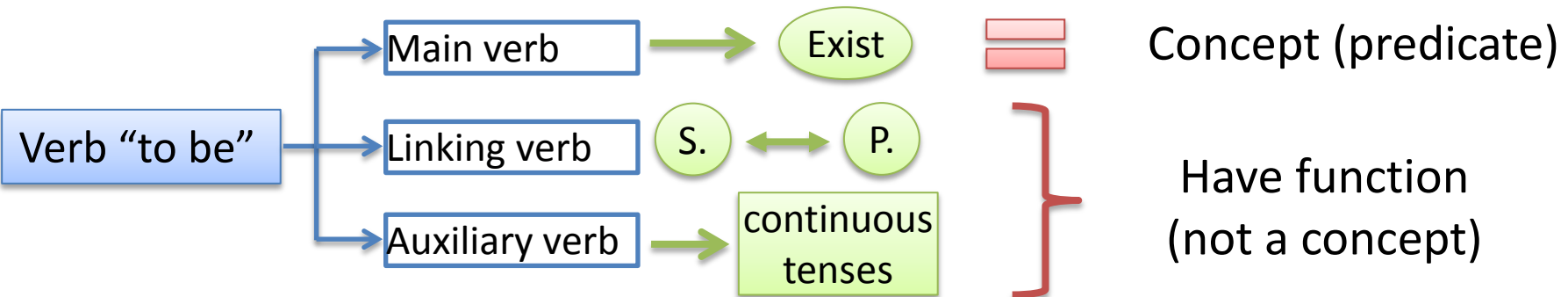
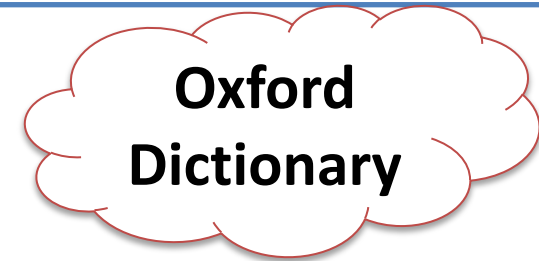
There once was a man

- [as copular verb] having the state, quality, identity, nature, role, etc., specified

The book was forty pages long

- Auxiliary verb

They are coming



➤ What is the nature of the verb “to be” in different languages??

In some languages, the verb “to be” as a copula verb is omitted in the present tense.

English	→	I am a human	} Zero copula
Russian	→	я — человек, <i>ya — chelovek</i>	
Hungarian	→	<i>ő ember</i>	
Arabic	→	أنا إنسان <i>?anā ?insān</i>	
Hebrew	→	אני אדם	

➤ Is verb “to be” the only verb of this kind ??

Look, Smell, Seem, Appear, Resemble, etc.

Example:

The verb “*appear*”



seem; give the impression of being:

She appeared not to know what was happening

Copula

come into sight; become visible or noticeable, especially without apparent cause:

The smoke appeared on the horizon

Main

Answer: The verb “to be” in the above sentence should not be represented as a UW, and “was” and “has been” are the same.

*The preposition "of" should be represented as a UW or as a relation?
(Consider "since", "from ... on", "in" or "during" instead of "of")*

- The words of, since, from and so on are prepositions which are classified as closed class (function words).
- Function words are not UWs.

Answer: The preposition “of” should be represented as a relation between “the most important novelist” and “the Victorian period”.

"generally regarded as" represented as UWs ("generally", "regarded", "as) or as an attribute (a downtoner, which lowers the truth effect of the declaration)?

Attributes may convey three different kinds of information:

- Information on the role of the node in the UNL graph (@entry).
- The information conveyed by bound morphemes and closed classes.
- Information on the (external) context of the utterance, i.e., non-verbal elements of communication, such as prosody, sentence and text structure, politeness, schemes, social deixis and speech acts.

➤ **What is the difference between ????**

He is the most important English novelist.

He is regarded as the most important English novelist.

He is generally the most important English novelist.

He is generally regarded as the most important English novelist.

1

He is the most important English novelist.



Truth- declaration

[S:1]

{org}

he is the most important English novelist

{/org}

{unl}

aoj(novelist:92.@def,he:70)

mod(novelist:92.@def,important:98.@most)



mod(novelist:92.@def,english:33)







{/unl}

[/S]

2

He is generally the most important English novelist.

generally
adverb UK  US  /ˈdʒen. ə r. ə l.i/ (US) /-ə-/

Definition       +1 1.3k

- **considering the whole of someone or something, and not just a particular part of them**
Your health is generally good, but you do have a few minor problems.
He wants more money to be given to the arts generally.
I shall now develop my previous point more generally (= to say more about what it includes).
- **usually, or in most situations**
The baby generally wakes up three times during the night.
*Well, generally **speaking** (= in most situations), it's quicker on public transport.*
- **by most people, or to most people**
It was generally believed at the time that both men were guilty.
The proposal has received a generally favourable reaction.

[S:2]
{org}
he is generally the most important English novelist
{/org}
{unl}
aoj(novelist:92.@def,he:70)
mod(novelist:92.@def,important:98.@most)
mod(novelist:92.@def,english:33)
man(important:98.@most,generally:64)
{/unl}

3

He is regarded as the most important English novelist.

regard
verb UK  US  /rɪ'gɑ:d/ (US) /-'gɑ:rd/ [T usually + adverb or preposition]

Definition        +1 1.3k

- **to consider or have an opinion about something or someone**
*Local people regard this idea of a motorway through their village **with** horror.*
Her parents always regarded her as the cleverest of their children.

[S:3]

{org}

he is regarded as the most important English novelist

{/org}

{unl}

mod(novelist:92.@def,important:98.@most)

mod(novelist:92.@def,english:33)

gol(regard as:11.@past,novelist:92.@def)

obj(regard as:11.@past,he:70.@topic)

{/unl}

[/S]

4

He is generally regarded as the most important English novelist.

2 different concepts

```
[S:4]
{org}
he is generally regarded as the most important English novelist
{/org}
{unl}
mod(novelist:92.@def,important:98.@most)
mod(novelist:92.@def,english:33)
gol(regard as:11.@past,novelist:92.@def)
obj(regard as:11.@past,he:70.@topic)
man(regard as:11.@past,generally:83)
{/unl}
[/S]
```


Results of your search



Your query was

generally regarded as

Generally regarded as

Here is a random selection of 50 solutions from the 138 found

138

[A19](#) 1356 Digital signal processing (DSP) is generally regarded as one of the fastest expanding areas of electronics.

[A7L](#) 781 And it's significant that a strand of films showing how a group of individuals from different social backgrounds could be brought together by war and learn to forget their divisions, generally regarded as the supreme achievement of wartime 'realism', was anticipated in Powell-Pressburger's *One of our Aircraft is Missing* (1941), in which a bomber crew bales out over Holland and works its way home with help from the Dutch resistance.

[AAP](#) 208 Olivier's performance in this role is generally regarded as somewhat flattering.

[ABS](#) 725 Such behaviour from an Amsterdam whore is generally regarded as a provocative come-on.

[AKR](#) 97 Mr Smith, who is generally regarded as the clear favourite for the leadership, suggested that the 'misrepresentation' of Labour's tax policy might well have caused some people not to support it.

[AR0](#) 1167 There, passenger traffic was light, and was generally regarded as a nuisance.

[ASY](#) 1056 A report on a pupil from within his own school will never be generally regarded as adequate, however conscientiously compiled.

[B1L](#) 1228 Millichip told them that Johnson had not breached any of the club rules and was generally regarded as one of their most loyal players.

[BIW](#) 647 most West-Pacific NICs, by structural adjustments in their economies and controlling real wage levels, have been able to cope with external economic shocks more easily than other countries — with more intra-trade and an increasing share of the world's fixed investment, economic prospects are generally regarded as sound by bankers;

[B25](#) 458 We will not go into great detail on the dangers of using statistics here but merely point to some precautions which must still be taken even when using statistics which are generally regarded as reliable.

[C95](#) 1152 *T. annectens* is generally regarded as the same species, although separation has been suggested.

[C96](#) 1047 If you have gained the impression that you are dealing with a balding manic-depressive axe-wielding lunatic, then I must hasten to add that I am generally regarded as a bit eccentric but quite harmless by my neighbours!

[C9E](#) 1090 If they are mentioned, it is because they have been recommended and are generally regarded as good courses.

[CED](#) 298 The results showed that Amasha was the least likely of the four to be treated with LGS and that Daeria, generally regarded as the most dangerous of the four, was the most likely reason for LGS usage.

[CET](#) 376 Neanderthals possessed bigger brains than we do; the Cro-Magnons, generally regarded as having wiped out the Neanderthal precursors, were similarly large-brained.

[CET](#) 999 Evolution is generally regarded as a slow process, but in fact it works remarkably swiftly.

[CLR](#) 792 After the fiasco of the EDC, the French National Assembly was generally regarded as being the place where the Treaty of Rome would stand or fall, particularly as the Fourth Republic was entering its final prolonged crisis, with ever weaker governments struggling to survive.

Results of your search



Regarded as

Your query was

regarded as

Here is a random selection of 50 solutions from the 4632 found.

4632

A0T 1284 For neurophysiologists and neuropsychologists, the way forward in understanding perception has been to correlate these dimensions of experience with, firstly, the material properties of the experienced object or event (usually regarded as the 'stimulus') and, secondly, the patterns of discharges in the sensory system.

A7W 264 Her promotion, following last week's appointment of Ms Margaret Beckett as shadow Treasury Chief Secretary, underlines Mr Kinnock's willingness to put women MPs into subject areas traditionally regarded as male preserves.

ABC 1467 Fish caught in gill-nets are generally regarded as being of poorer quality, because they are often dead for many hours before coming aboard the fishing vessel.

AJV 861 A novel like *The Middle Age* of Mrs Eliot, for instance, could be regarded as a latter-day *Middlemarch*, with Meg rising up and breaking free as Dorothea does, and it is possible to detect the influence of Wilson's favourite writer, Dickens, in works as dissimilar as *Anglo-Saxon Attitudes* and *No Laughing Matter*.

AMM 194 For many years they were regarded as colonial coelenterates, but it is now certain that they are unrelated to the jellyfish and their allies, and in fact are distant cousins to a small group of tube-dwelling organisms with little fossil record, which belong to the minor phylum Hemichordata.

APS 722 Flaubert's use of style indirect libre in order to voice the consciousness of his protagonists, is regarded as the forerunner of the stream of consciousness technique popularized by Joyce and other twentieth-century modernists.

ASC 1297 As she was also trying out several other would-be choreographers and having new ballets made for Sadler's Wells by Ashton and Howard too, she could hardly have done more for someone who was regarded as promising (by some people as highly promising) but unproven.

ATI 1497 I think he was regarded as a bit of a failure in Ukrainian circles because I didn't speak the language or dance.'

B2P 1059 When someone has been asked to make over on death whatever remains of an inheritance, and from the price of objects sold buys other objects, he is not regarded as having diminished [the estate] in respect of the objects sold... but the objects thus bought should be made over in place of the ownership which has changed....

B74 800 Nor can the stability argument be regarded as the last word, as we do not know whether general relativity itself will be valid right up to the space-time singularity.

BM2 1094 No matter how ruthless and cruel the evolutionary process may appear to be, it nevertheless must be regarded as having been inevitably so.

BMK 215 These are regarded as 'transitional substances' that could help wean the world off its dependence on CFCs, but since they have their ozone-depleting potential (albeit low) the protocol will probably be extended to ban them early next century.

CCU 764 11.1 During the continuance of this Agreement the Proprietor shall not without the consent in writing of the Publisher prepare or edit for any other publisher any work that is an expansion abridgement or revision of the Work or of any part of it or publish or cause to be published any work on the same subject which may reasonably be regarded as likely to conflict with the sale of the Work.

CG6 624 This line of reasoning then raises the question of what might be regarded as an accurate description of the language of ordinary children at different ages.

CMN 517 Rather than being regarded as actors who make their own history, individuals are to be seen as the 'supports' of social practices who maintain and reproduce them.

Results of your search



generally

Your query was

generally

11456

Here is a random selection of 50 solutions from the 11456 found.

[A0M 108](#) Generally, it is best to send out a strong opener, that is someone who can be relied upon to go all out for a victory.

[A0M 1155](#) The aggressive fighter is often weak on defence and generally prefers just to smother your attack so that he can proceed with his own.

[A7L 65](#) Such thinking has generally been wrong on every count.

[AB9 1925](#) 'Oh come now, Bill, Oxford senior common rooms are full of people quite as untidy and generally as odd as Gerhard.

[ACJ 918](#) This survey has shown that, generally speaking, the existing range of offences seems to emphasize the result, the degree of foresight and the status of the victim as the critical issues in grading crimes of physical violation.

[AD0 241](#) Expect that generally things will change for the better.

[B10 1059](#) Sometimes (but rarely) you punish his unacceptable behaviour, generally when there are visitors.

[B15 47](#) Things generally are much more complicated and confusing than they need be.

[B3B 680](#) Students generally will be examined on the taught elements of the courses by cumulative assessment or by written examinations, or a combination of the two.

[B73 842](#) Such heavy particles can not be created at even the highest energies particle accelerators can reach, but they could have been produced copiously in the aftermath of the big bang, with which, cosmologists generally believe, the Universe began.

[B7E 287](#) A Rumanian doctor, Ana Aslan, who was using these injections, noticed that her elderly patients appeared much more lively, and generally healthier after long-term procaine injections.

[BMC 543](#) Performances are idiomatic, with well sprung rhythms and sensibly chosen tempi, and benefit from a generally cool and restrained approach.

[CBB 580](#) In 1874 the Mercury reported that the custom was in decline but 'formerly several gangs of 'plough-witches', ludicrously dressed and with hideous painted faces, came in from the country and visited the tradesmen's shops, levying "black-mail" and generally terminating their begging rounds with drunken orgies'.

[CBU 1778](#) But until very recently systems were generally aimed at the top end of the market, and the cost was high enough to make acquisition a major undertaking.

[CBY 2910](#) Reputation of the jurisdiction generally.

[CE2 308](#) Therefore a passenger jumping off a moving bus and injuring himself would bring the motor vehicle within this section, but not generally where a driver has left his stationary motor vehicle parked on a road.

[CMB 488](#) While this is a continuing habit of government it is worth pondering whether governments have become more adept at doing this and whether they have encouraged the growth of a machinery with which to co-ordinate media policy generally.

- Generally regarded as is a collocation composed of two different concepts.
- We have to be consistent in our decision:

If we considered “generally regarded as” an attribute
This may make a mess and subjective judgment;
anyone can consider anything to be an attribute.

Answer: “generally regarded as” should be represented as two distinct UWs.

*"most" should be represented as a UW or as a superlative marker
(attribute assigned to "important"?
(Consider "greatest" instead of "most important")*

➤ **What is the type of the adverb "most" ??**

- Degree or Quantity adverb **Ex:** very- most

➤ **Degree and quantity adverbs in different languages:**

- In many languages, degree and quantity adverbs are formed by using a definite article and suffix attached to the adjective to add the **Superlative Feature** to the words.

English Example

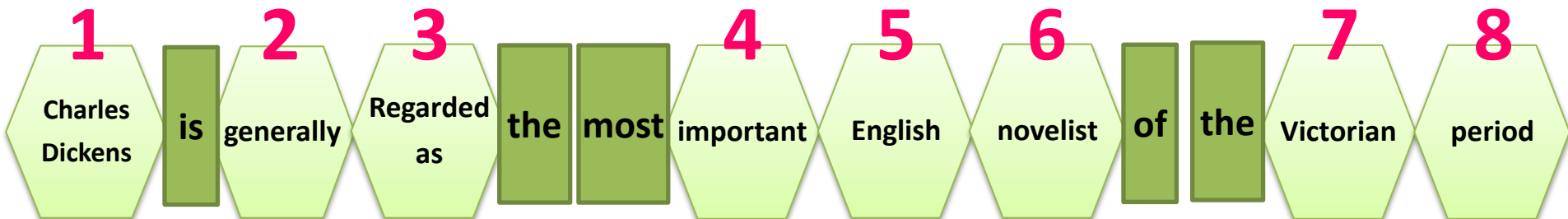
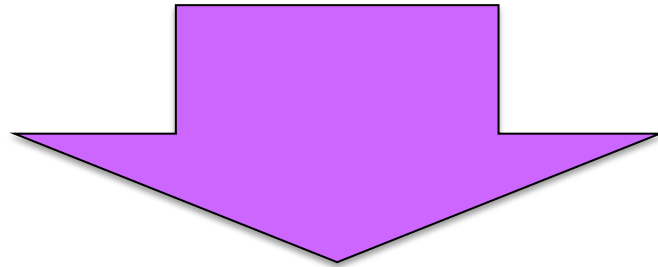
Most beautiful = Tallest

Answer: The adverb "most" in the above sentence should not be represented as a UW.

The superlative form of adverbs is language-dependent

Now, we can count how many UWs in this sentence:

Charles Dickens is generally regarded as the most important English novelist of the Victorian period



8 UWs

Question 2:

"Charles Dickens" should be represented as a permanent UW or as a temporary UW?

- The UNL Dictionary contains only permanent UWs.
- URLs, e-mail addresses, phone numbers, formulae etc. are untranslatable expressions (transliteratable) and are not included in the dictionary, but may be used in the UNL graphs as temporary UWs.

Consider

- “Charles J Dickens” **person name**, Charles Dickens **Museum**, Charles Dickens **restaurant**, Charles Dickens **School** and other entities named “Charles Dickens”.
- **The size** (and the maintenance) of the **UNL Dictionary**, **in case** you suggest to **treat them all as permanent UWs**.

To answer this question:

- ✓ Different theories about the meaning of proper nouns.
- ✓ The boundaries between common and proper nouns.
- ✓ Proper nouns in the mental lexicon.
- ✓ Proper nouns in NLP.
 - Document retrieval.
 - Question-answering.
 - Named entity recognition.
 - Agreement and accuracy.

Do Proper Nouns have Meaning?

The Descriptivist Theory of names

- **Meanings** (semantic contents) **of names** are identical to the **descriptions associated with them by speakers**, while their referents are determined to be **the objects that satisfy these descriptions**.

The Causal Theory of Reference

- A name's referent is fixed by an original act of naming where upon the name becomes a rigid designator of that object.
- Later uses of the name succeed in referring to the referent by being linked to that original act via a causal chain.

Discussion on the meaning of proper nouns is endless...

The Description Theory of the reference of proper names:

Given any subject S, name 'stan', and family or cluster of properties ϕ such that S believes 'Stan has ϕ ', an utterance of 'Stan' by S denotes whatever unique thing has most of the ϕ s.

Consequences of the theory:

(1) semantic

- (a) the individual S denotes by 'Stan' has most of the ϕ s
- (b) whatever unique individual has most of the ϕ s is the individual S denotes by 'Stan'

(2) epistemological

- (a) S knows *a priori* that if Stan exists, Stan has most of the ϕ s
- (b) S knows *a priori* that whatever has most of the ϕ s is Stan

(3) metaphysical

- (a) Necessarily, if Stan exists, Stan has most of the ϕ s
- (b) Necessarily, whatever has most of the ϕ s is Stan

Kripke-style objections

The man I refer to as 'Thales' or 'Jonah' (e.g.) may have hardly any of the properties I think he has

1. Often our descriptions are not complete enough to pick out anyone

uniquely

- example: entry on Ramses VIII: "one of several Egyptian Pharaohs about whom nothing is known."
- Twin Earth examples seem to show that this could be the *usual case*

2. In other cases it seems our descriptions might pick out the *wrong person*

it seems that we could discover that (e.g.) Thales didn't maintain that all is water, wasn't Greek, etc.

This is clearly false. Even if we are *given* that Thales *did* do the things we believe he did, he surely *might* have failed to. (If 3a were true, 'Stan' would not

Descriptionist responses (cf. Searle 1982)

Maybe. But we could not be *wholly* mistaken. We will be correct, for example, in thinking that 'Thales' when we use it denotes whatever was denoted by the person from whom we got the name. (In cases where we have named an object, we will be correct in thinking that the name denotes whatever we dubbed when we bestowed the name.) We should think of the denotation, for S, of 'Stan' as whatever has a weighted most of the ϕ s.

We have a uniquely determining description for Ramses VIII: the Egyptian Pharaoh between Ramses VII and Ramses IX.

Even in Twin Earth cases there will be a description which picks out the referent but not its Twin Earth counterpart (e.g. 'person I am seeing now').

Maybe if we just counted ϕ s. But if we weight the importance of the ϕ s the answer will always come out right.

Granted. But the description theory need not have consequence (3): we might mean by 'Stan' merely "the person who *actually* has most of the ϕ s." (Notice that this is a rigid designator.)

Some replies on behalf of the "new" or "causal" theory

These descriptions will be correct, but only because they reproduce the history of S's acquisition of the name. (Weighting the ϕ s will work only if we give weight 1 to the description "whatever I dubbed when I bestowed the name," in the case of names, and to "whatever the person I got the name from referred to," in the case of others; and give weight 0 to all other descriptions.)

This is true only if 'Ramses VIII' is not really a proper name but merely short for e.g. the description "the seventh Egyptian pharaoh named 'Ramses'." If 'Ramses VIII' is a genuine proper name, the description could turn out to be false. (Maybe they wanted to reserve that name for a special occasion.)

This *only* works if your description contains an indexical expression (e.g. 'I', 'now', 'here', 'this'). But then we need to address the question whether *indexical* reference is determined by description, and here the answer seems clearly 'no'. (All the non-indexical descriptions I have of myself apply equally to my Twin Earth counterpart.)

Only if the weighting matches the "new

<http://www.trinity.edu/cbrown/language/>

<http://www.trinity.edu/cbrown/language/descriptiontheoryofnames.pdf>

1. The boundaries between common and proper nouns

1.1. In function

- In some cases, the semantic function of proper nouns and common nouns coincide...

“The president approached his interlocutor”



“Obama approached his interlocutor”

“She took two Tylenol and went to bed”



“She took two acetaminophen tablets and went to bed”

1.The boundaries between common and proper nouns

1.2. In meaning

Some proper nouns acquire common properties:

Khlestakov

A fictional character



Common meaning:

unrestrained boasting, lying

“khlestakovshchina”



Common noun

سي السيد (Si Elsayed)

A fictional character



Common meaning:

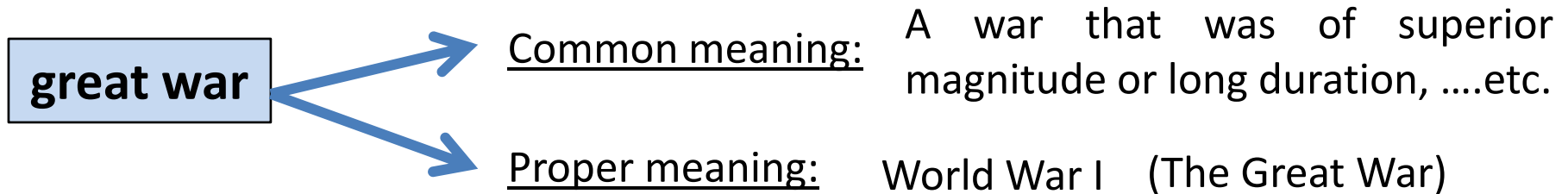
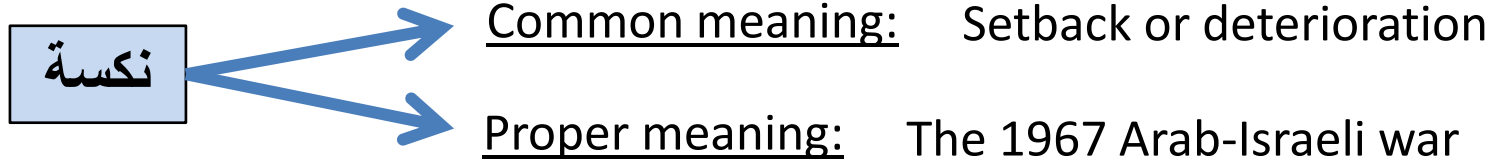
a controlling and demanding husband

1. The boundaries between common and proper nouns

1.2. In meaning

Some common nouns acquire proper properties:

Examples:



2. Proper Nouns and the Mental Lexicon

- Proper nouns assume a great deal of common knowledge between speaker/writer and listener/reader.
- Proper nouns show different degrees of accessibility and their interpretation requires different types of knowledge.

If the speaker's assumption about the listener's knowledge is false,

AND

If the knowledge conveyed by the speaker in the conversation does not allow the listener to bypass the knowledge deficit

Speaker

I went to Charles Dickens

I went to Charles Dickens restaurant

??????

Listener



The proper noun automatically becomes a source of trouble for the continuing communication.

3. Usefulness in Information Retrieval and Extraction

3.1. Document Retrieval

- Knowledge on proper nouns can be very effective for document retrieval

The query: "Les troubles politiques et civils au Sénégal en 1993"
(Political and civil unrest in Senegal in 1993)

Event

can have many variations (synonym and structure).

Time

restricts documents on a temporal window.

Place

more likely to be present under this form in the documents retrieved.

3. Usefulness in Information Retrieval and Extraction

3.2. Question-Answering

There are two main uses of a proper noun thesaurus in question-answering:

- ✓ Verifying if a potential answer corresponds to what is needed.

Example: “what is the fastest animal?”

A prancing black horse on a yellow background.

✗ Ferrari

A car brand

✓ cheetah

An animal

- ✓ Answering questions.

Example: “what is a Ferrari?”

✓ A Ferrari is a car

3. Usefulness in Information Retrieval and Extraction

3.3. Named-entity Recognition

- Finding proper nouns in dictionaries would greatly aid named-entity recognition applications.

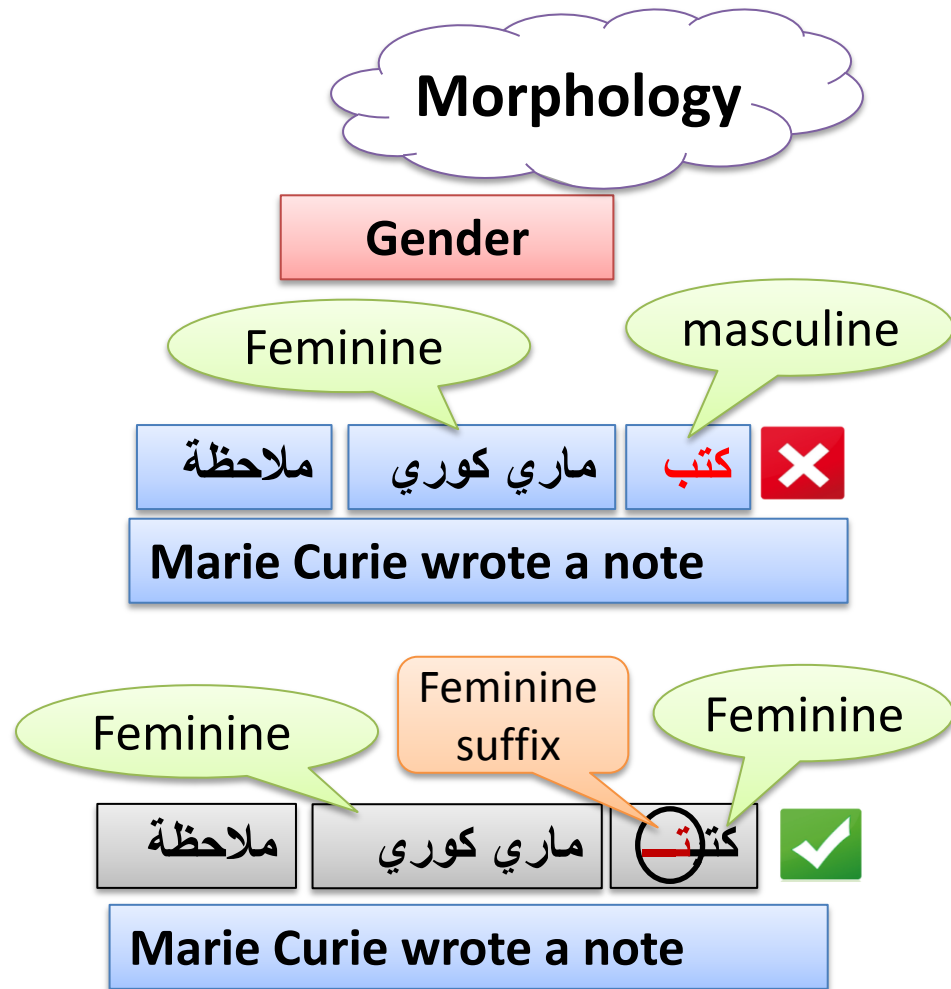
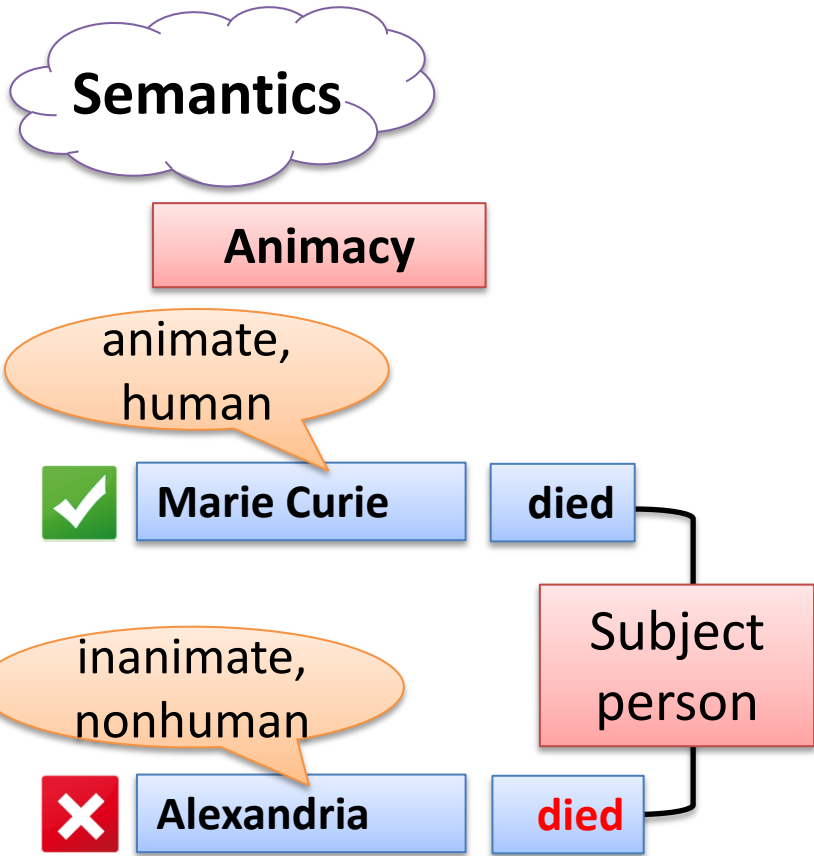
Sheikh Mohamed Naseif was born on the 18th of Ramadan 1302 a.h.

[Sheikh] "" (ANI, TTL, PPN);
[Mohamed Naseif] "" (ANI, HUM, PPN);
[18] "" (DAY, PPN);
[Ramadan] "" (MTH, PPN);
[1302] "" (YR, PPN);

Sheikh <PPN, TTL> Mohamed Naseif <PPN, HUM> was born on the
18th <PPN, DAY> of Ramadan <PPN, MTH> 1302 <PPN, MTH> a.h.

4. Agreement and Accuracy in the Target Language

Proper nouns are necessary for achieving morphological agreement and semantic validity in the target language.



The answer is:

- Assuming that proper nouns are temporary UWs inevitably implies ignoring some aspects of their meaning.
- Storing proper nouns as permanent UWs ensures transferring the deep meaning of the original text to the target language, which is the goal of the UNL

We choose to store proper nouns as permanent UWs in the UNL dictionary.

A challenge !!

But: Shall we store all proper nouns?

If we were to classify proper nouns into proper nouns **worth storing** and not so important proper nouns.

How will we draw this line?

Significance varies according to **place**:

4th of July

→ **USA:** The country's Independence Day

→ **Outside the USA:** the fourth day in July

Significance varies by **time**:

Nader Bakkar

→ **2 years ago:** One of the personal names

→ **Now:** The spokesman of Al Nour Party and a household name

Thus, we choose to store **all** proper nouns as permanent UWs in the UNL dictionaries

Maintaining the UW dictionary

- The dictionary size will be, of course, too overwhelming for the system to function efficiently.
- Thus, we suggest that the proper noun thesaurus be a stand-alone lexicon to be referred to only when needed.
- Since the number of proper nouns is countless, and storing them is an endless task...

We thought of giving priority to the most frequent proper nouns in a representative corpus, so that the most frequent are defined first, and so on.

Question 3:

"hunger" (= "a physiological need for food"), "hungry" (= "feeling hunger"), "hungrily" (= "in the manner of someone who is very hungry") and "hunger" (= "to cause to experience hunger") should be represented as simple, compound or complex UWs?

Lexical Item (English)	Simple UW	Compound UW	Complex UW
hunger	hunger	hungry.@ness	a physiological need for food
hungry	hungry	hunger.@full_of	feeling hunger
hungrily	hungrily	hunger.@full_of.@manner hungry.@manner	in the manner of someone who is very hungry
hunger	hunger	hunger.@full_of.@make hungry.@make	to cause to experience hunger

Consider:

- ✓ Some of these concepts are not lexicalized in all languages.
- ✓ The actual importance of part-of-speech in lexical semantics.
- ✓ The actual "compositionality" of these concepts.

To answer this question:

- ✓ Lexicalization and lexical gaps.
- ✓ Parts of speech in different languages
- ✓ Compositionality of words.

Simple UWs

Each of the lexical items “hunger”, “hungry” or “hungrily” which correspond to different parts of speech should be represented as isolated nodes in the UNL graph.

Parts of Speech in different languages:

- Currently, Linguists recognize that the list of the eight word classes is drastically simplified and artificial, and they are not applicable to all languages (*Zwicky, Arnold, 2006*)
- Many languages do not have a distinction between different parts of speech.

Example:

- In Thai, there is no morphological distinction between adjectives and adverbs.
- In Japanese, in some cases, there is no distinction between adjectives and verbs.

- Some language do not even have a name for some of the parts of speech.

Part of Speech	Greek	Thai	Welsh	Swahili	Filipino	Croatian
noun	---	นาม	enw	---	pangngalang pambalana	imenica
verb	ρῆμα	คำกริยา	berf	kitenzi	pandiwa	glagol
adjective	επίθετο	คำคุณศัพท์	ansoddair	kivumishi	pang-uri	pridjev
adverb	επίρρημα	---	---	kielezi	---	prilog
conjunction	συνδυασμό	ร่วม	ar y cyd	kushirikiana	kasabay	zajedno
preposition	---	บุพบท	---	Nomino	pang-ukol	---
article	άρθρο	บทความ	Erthygl	makala	artikulo	članak
pronoun	αντωνυμία	คำสรรพนาม	Rhagenw	---	panghalip	zamjenica
interjection	επιφώνημα	การพูดสอดขึ้น	ymyriad	interjection	pagsabad	ubacivanje

- Depending on the part of speech in defining a UW gives rise to a **problem** regarding **lexicalization and lexical gaps**.

Lexical gaps

When a lexicalized concept in one language does not have a corresponding lexical unit in another language.

Example

Romanian lacks a word to represent the concept "**shallow**".

Shallow



puțin adânci

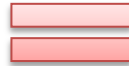


Not so deep

Using simple UWs

Simple word-to-word replacement can lead to semantic problems.

The English adjective **"Hungry"**



The French adjective **"affamé"**

Example:

I am hungry.

Inappropriate French translation

Je suis affamé.

Unacceptable



Appropriate French translation

J'ai faim.

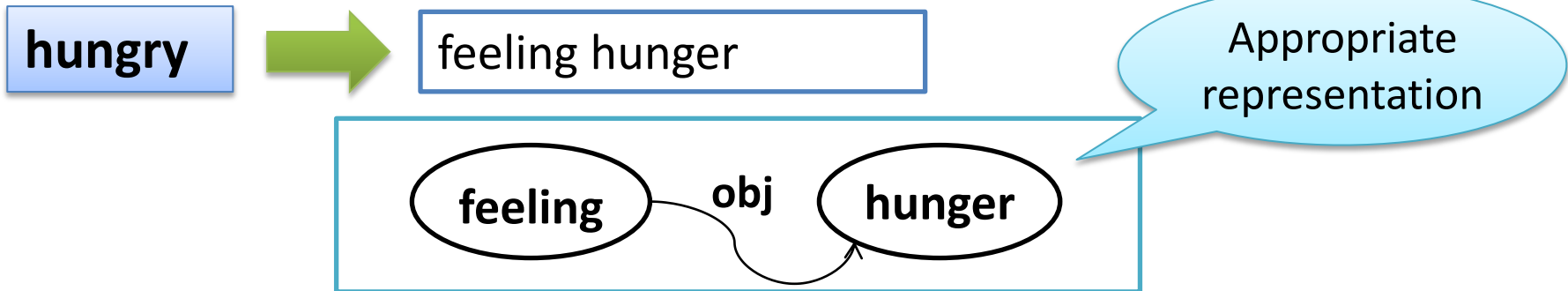
Acceptable



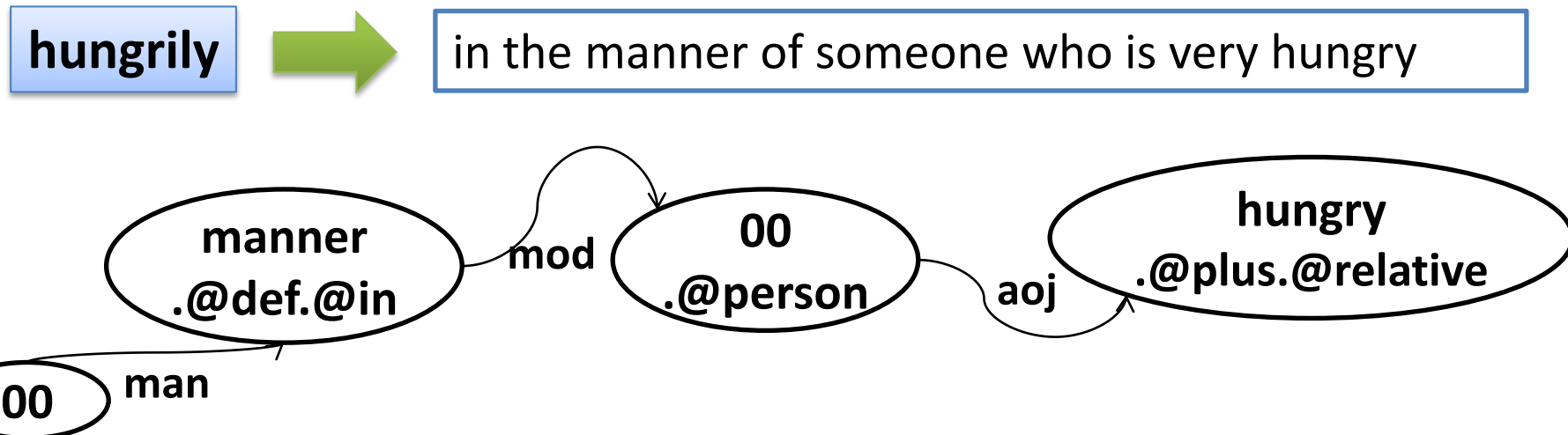
Simple UWs in this case will be language-dependent and would not be flexible enough to be universal.

Complex UWs

Each of the lexical items “hunger”, “hungry” and “hungrily” should be represented as a sub graph in the UNL graph.




Is this representation always appropriate???



Example: He pounced on the food hungrily



```
[S:1]
{org}
He pounced on the food hungrily
{/org}
{unl}
agt(pounce:82.@past,He:61.@male.@3)
obj(pounce:82.@past,food:65.@on.@def)
man(pounce:82.@past,hungrily:35)
{/unl}
[/S]
```



```
{org}
In the manner of someone who is very hungry
{/org}
{unl}
man(00:66,manner:14.@def.@in)
mod(manner:14.@def.@in,someone:51.@person)
aoj(hungry:94.@plus.@relative,someone:51.@person)
{/unl}
```

Difficulty of processing

Creating sub graphs would make this proposal almost unfeasible.

- ✓ However, this representation will be useful in case the concept is not lexicalized at all in a certain language.

Shallow



puțin adânci



Not so deep

Complex UWs may be misleading and too difficult to process.

Compound UWs

Compound UWs assume that a single lexical item such as “hungry” or “hungrily” can be reduced into a **smaller lexical part + an attribute**.

This hypothesis is supported by the principle of compositionality.

Compositionality

A single concept can be reduced to smaller lexical parts.

Smaller lexical part



morpheme



carrying the most essential aspects of semantic content



Carrying secondary aspects of meaning

hungry



ly



hungrily



(New Concept)

The combination of a smaller lexical part + an attribute can be considered as the realization of derivation.

Derivation

root

The primary lexical unit which carries the most essential aspects of semantic content.

affix

A morpheme that is attached to a word root to form a new word.

Representing the adjective “hungry” as a compound UW

English

hungry

UW

hunger.@full_of

French

ai faim

Each language will be able to represent the meaning conveyed in the UW in the way it finds acceptable.

The answer is :

The compound UW is the most appropriate representation of meaning since:

- > it overcomes the problem of lexicalization posed by non-corresponding parts of speech in different languages, thus;
- > it offers a greater chance to achieve universality;
- > it does not place any restrictions to processing.

Question 4:

Antonyms (“mortal” and “immortal”, “hot” and “cold”, and “son” and “father”) should be represented as a single UW or as different UWs?

Consider

- ✓ Absolute opposites such as "mortal" x "immortal" (which can be inferred by an attribute such as @not).
- ✓ Gradable opposites such as "hot" and "cold" (which require: hot.@extra, hot.@plus, hot, hot.@minus, hot.@not, hot.@not.@minus, hot.@not.@plus and hot.@not.@extra)
- ✓ Relational opposites such as the converse "son" and "father" (which require @converse).

Antonyms

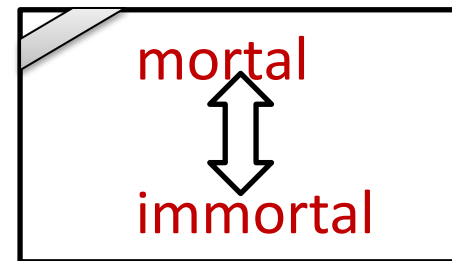
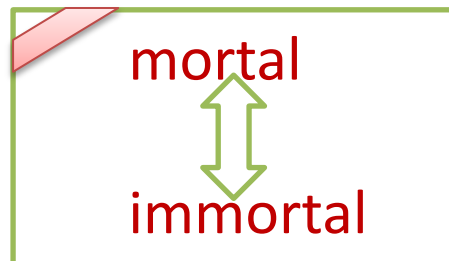
Complementary pairs

Single ↔ Married

Mortal ↔ Immortal

- The assertion of one entails the negation of the other.

* Joking aside, you can't really be "a little pregnant" Or "kind of married."



Same concept of mortality

Same concept of immortality


Which one is non-marked?

Determining the non-marked

1


The how test


“how long” or “how deep” 

“how short” or “how shallow”. 

2


The quantification test

“3 feet long” or “15 meters high” 

“3 feet short” or “15 meters” low 

3

Nouns derived from adjectives

length – depth. 

shorth – shallowth. 

Non-marked

long

short

long@not

=

short

How about other complementary antonyms ?

Input ↔ Output

Push ↔ Pull

Exhale ↔ Inhale

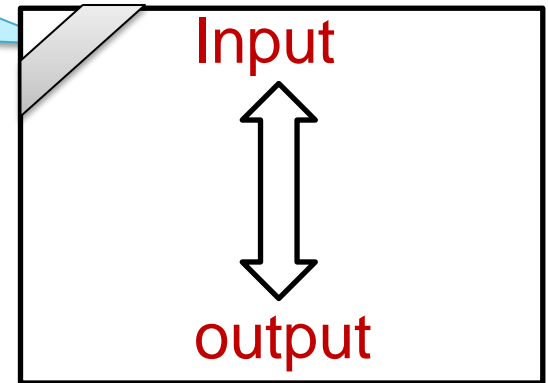
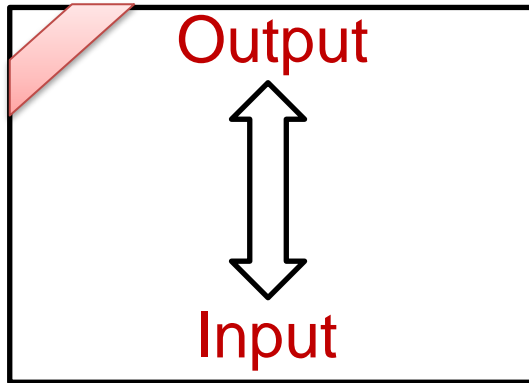
Dead ↔ Alive

Married ↔ Single

Right ↔ Wrong

Can we determine the marked and the non-marked using these tests?

Which one is
non-marked??



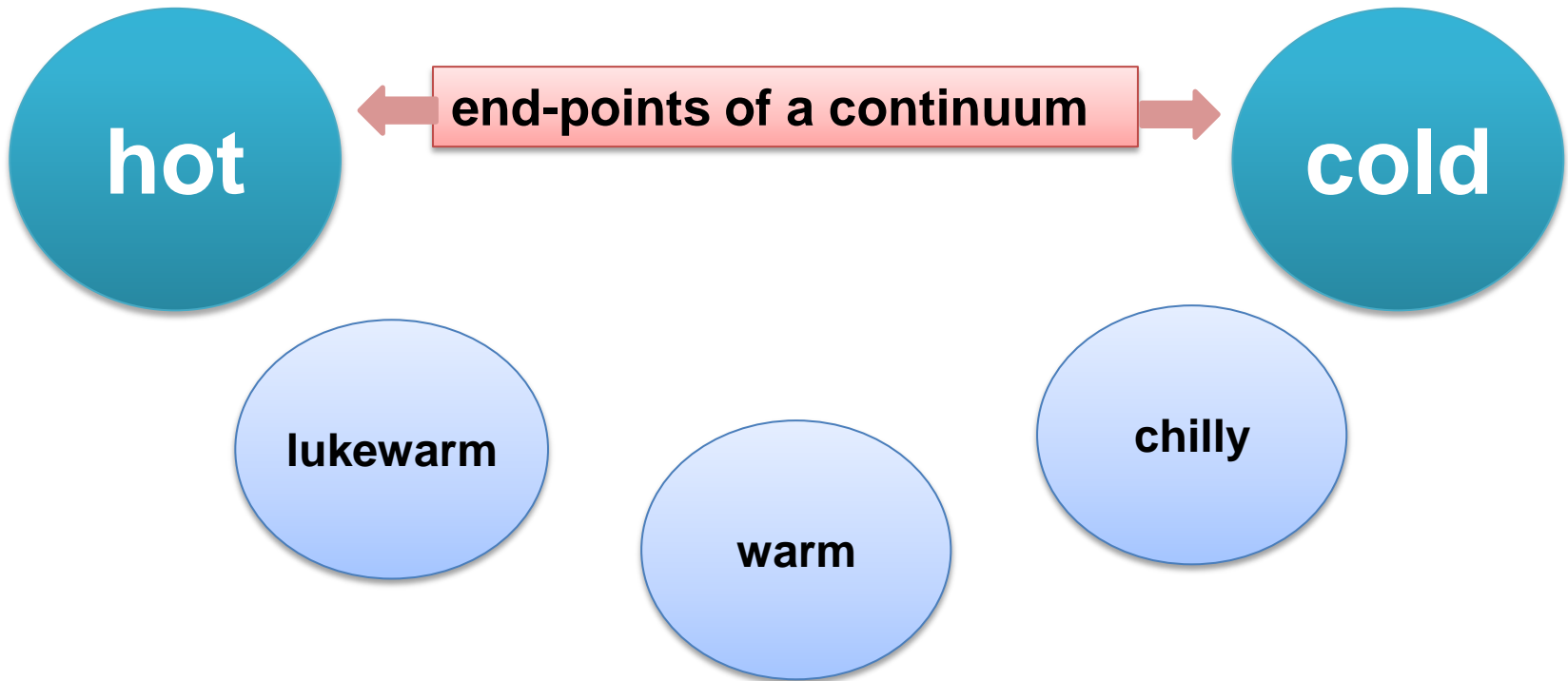
output@not

or

input@not

Complementary antonyms can only
be represented as **single UWs**

Gradable antonyms



Does not
equal "cold"

hot

cold

Not an
either-or
relation

hot@not

≠

cold

A more-
less
relation

hot

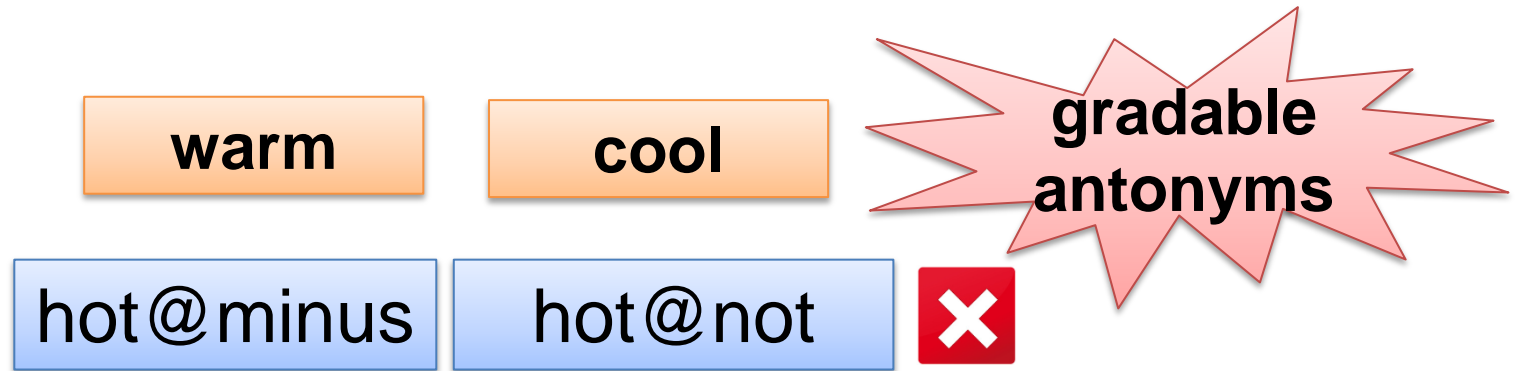
warm

chilly

cold

cool

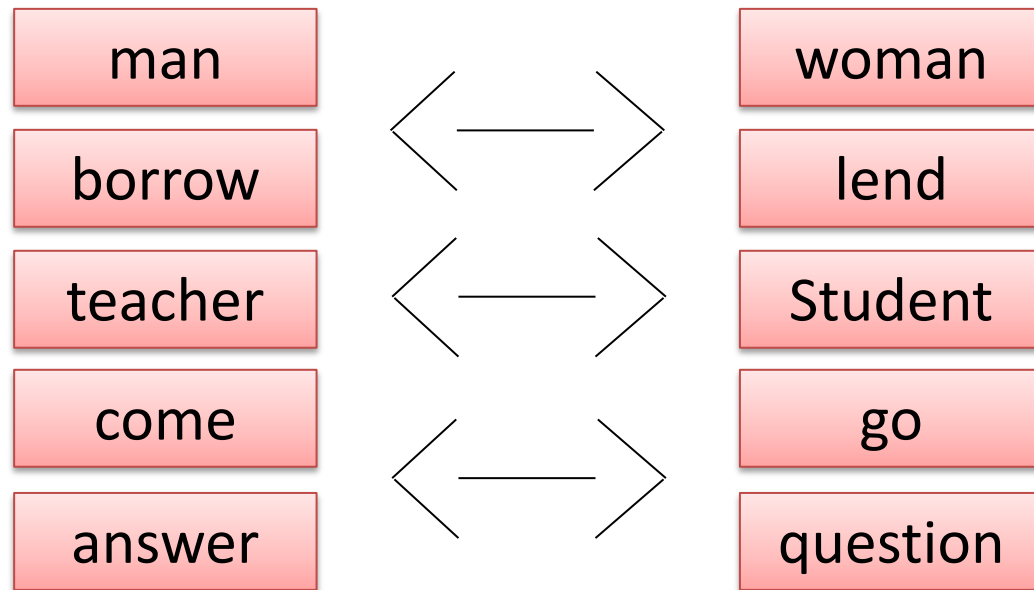
✓ Grades are also antonyms.



Warm and **cool** are in the same distance from the middle of the scale but shorter than hot and cold.

Relational antonyms

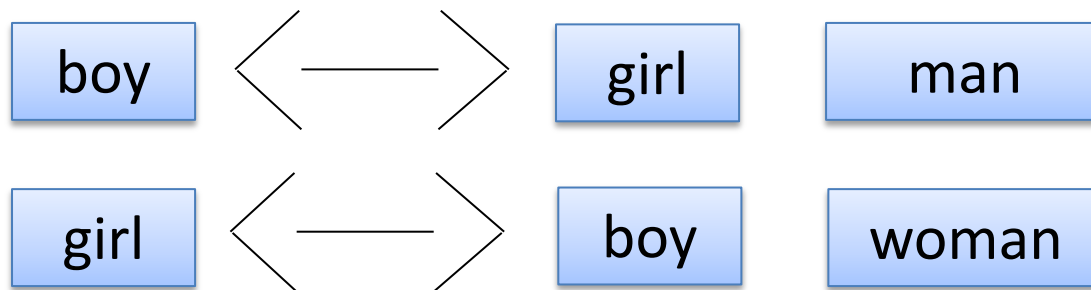
- Pairs of words displaying a symmetrical relation in their meaning
- One member of a relational antonymous pair depends on the existence of the other.



- Some concepts have more than one antonym.



- Each member is the opposite of two other members.

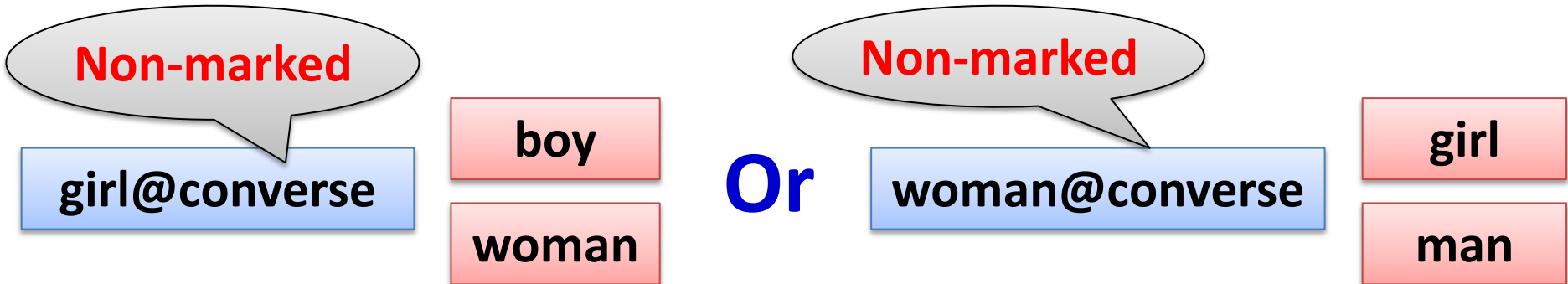


- It is difficult to predict all possible relational antonyms.

Which one is the **antonym** of the word “boy”????



If we can decide the antonym, which pair is **non-marked**???



Relational antonyms cannot be represented as **compound UWs**

The answer is :

It is difficult to determine which one is non marked?

Antonyms cannot be represented as: single
UWs, but as distinct UWs

Question 5:

"Farbfernsehgerät" ("color television set", in German) should be represented as a simple or complex UW?

- Every concept lexicalized in at least one language must be defined as a permanent UW.
- "Farbfernsehgerät" is considered to be lexicalized in German (found in German dictionaries as one single entry). The English equivalent "color television set" not lexicalized in English (not found in English dictionaries).

Consider

- ✓ "Farbfernsehgerät" is formed by "Farbe", "Fernsehen" and "Gerät" (not simple compound of three words, but underwent spelling changes).
- ✓ The case of compounds such as "baby-talk", "bittersweet" and "skinhead".
- ✓ The case of idioms, such as "all ears", "closed book" and "cold feet".

“Farbfernsehgerät” is represented as one single independent entry in dictionaries.

The screenshot shows a web-based dictionary interface. At the top, there are tabs for 'Dictionary De - En', 'De - Es', and 'De - Pt', along with a 'Vocabulary trainer' link. Below the tabs are navigation buttons for 'Spec. subjects', 'Grammar', 'Abbreviations', 'Random search', and 'Preferences'. A search bar is set to 'De-En Dictionary' with a 'tolerant' search mode and 'w/o phonetic' options. The search term 'Farbfernsehgerät' is entered and highlighted with a red speech bubble. A 'search' button is next to it. To the right, a dropdown menu lists related terms: 'Farbfehlsichtigenmodus', 'Farbfehlsichtigkeit', 'Farbfernsehen', 'Farbfernseher', 'Farbfernsehgerät', 'Farbfilm', 'Farbfilmverfahren', 'Farbfleck', and 'Farbfoto'. Below the search bar, it shows '1 result for Farbfernsehgerät' and 'Word division: Farb-fern-seh-ge-rät'. The results are split into two columns: 'German' and 'English'. The German column lists 'Farbfernseher {m}; Farbfernsehgerät {n}' and 'Farbfernseher {pl}; Farbfernsehgeräte {pl}'. The English column lists 'color TV; colour TV set [Br.]; color television set [Am.]' and 'color TVs; colour TV sets'. At the bottom, there are links for 'Example sentences', 'Synonyms', 'Proverbs, aphorisms, quotations', 'Wikipedia', and 'Google: Web search'. A footer note states 'No guarantee of accuracy or completeness!' and '©TU Chemnitz, 2006-2012'.

Dictionary De - En De - Es De - Pt Vocabulary trainer

Spec. subjects Grammar Abbreviations Random search Preferences

Search in De-En Dictionary tolerant w/o phonetic

Search for: ä ö ü ß search clear

▲ Farbfehlsichtigenmodus
Farbfehlsichtigkeit
Farbfernsehen
Farbfernseher
Farbfernsehgerät
Farbfilm
Farbfilmverfahren
Farbfleck
Farbfoto
▼

1 result for **Farbfernsehgerät**
Word division: Farb-fern-seh-ge-rät

German **English**

Farbfernseher {m}; **Farbfernsehgerät** {n}
Farbfernseher {pl}; **Farbfernsehgeräte** {pl}

color TV; colour TV set [Br.]; color television set [Am.]
color TVs; colour TV sets

Search further for "Farbfernsehgerät":
 [Example sentences](#) | [Synonyms](#) | [Proverbs, aphorisms, quotations](#) | [Wikipedia](#) | [Google: Web search](#)

No guarantee of accuracy or completeness!
©TU Chemnitz, 2006-2012

✓ "Farbfernsehgerät" composed of "Farbe", "Fernsehen" and "Gerät".

Similar cases in other languages:

Concept components are not important:

English:

[Stagflation] = Stagnation + Inflation

Arabic:

[ما فيش] = ما + في + شيء
nothing = no + in + thing

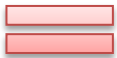
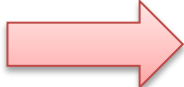
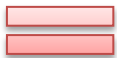

New concept lexicalized and conceptualized in the language

- ✓ "Farbfernsehgerät" (color television set) is not lexicalized yet in English.

Concept Complexity and simplicity

- It is not important for the word to be lexicalized in other languages to decide whether the word is simple or complex.
- English language cannot be used as the standard for deciding whether a word is simple or complex.

Basque language

<i>janaria prestatu</i>		Literally: 'to prepare food'		To cook
<i>bizikletan ibili</i>		Literally: 'to walk on a bicycle'		To bike

Complex in Basque but simple in English.

Challenges :

English

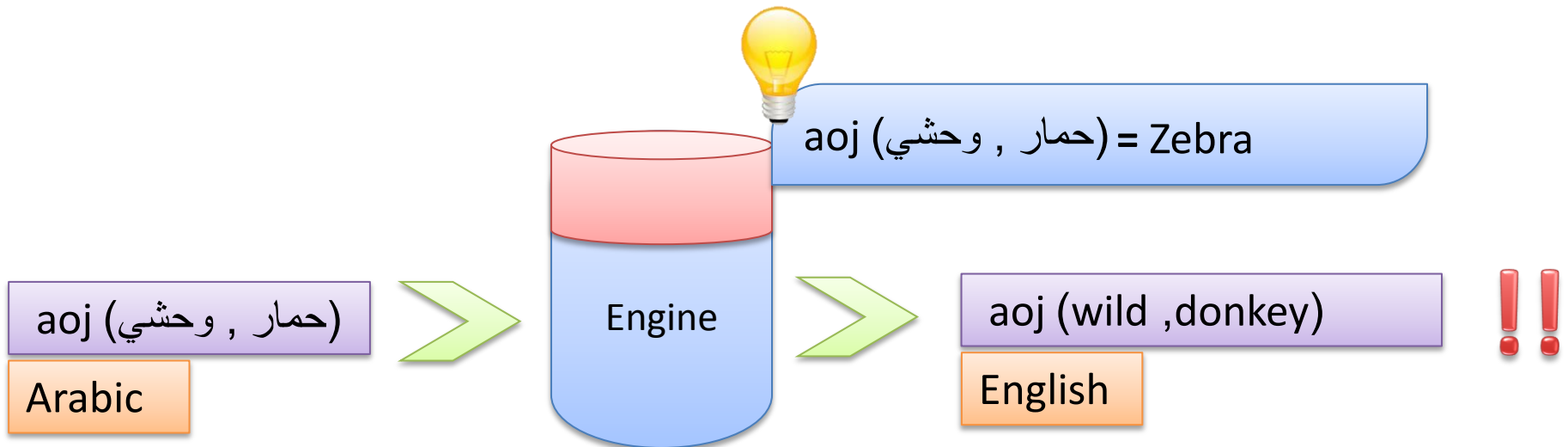
hare  One simple UW  Wild rabbit

Arabic

أرنب بري  Two simple UWs  أرنب  بري

Challenge solution

- Developing a smart tool that can predict and suggest that أرنب بري can be better translated into Hare in English rather than Wild Rabbit, as Hare has a higher priority.
- The tool should also suggest translating حمار وحشي into zebra not wild donkey.



Considering Multiword Expressions:

- Unpredictable multiword expressions will be represented as **simple UWs**.

Idioms

Closed book, cold feet, all ears.

Phrasal verbs

Put up with

Unpredictable compounds

Skinhead, zebra crossing, cloud nine.

The answer is:

“Farbfernsehgerät” should be represented as a simple UW based on the general principles of lexicalization and conceptualization.

Summary and Final Conclusion

In creating UWs we have to take the following points into consideration:

- Not all open lexical categories should be represented as UWs.
- Verb “to be” when used as a copula verb should not be represented as a UW.
- Deciding whether a multiword expression is a single UW or more depends on the conceptualization, compositionality and computational predictability.
- All prepositions have function and should be represented as Relations or Attributes.
- Although degree and quantity adverbs are open lexical categories, they should not be represented as UWs because they only add a quantitative feature.

- It is important to store all proper nouns in the UNL dictionary.
- Compound UWs is the best way to fill lexical gaps.
- Complex UWs can be useful in some cases, but in general can be difficult to process.
- Antonyms should be represented as different UWs because:
 - We cannot decide which one of the pair is non-marked.
 - The grades in gradable antonyms themselves have antonyms.
 - Relational antonyms may have more than one opposite.
- If a word is lexicalized and conceptualized in a language, it should be represented as simple UW even if not lexicalized in other languages.



Thank you