Outline

- Linguistic Theories of semantic representation
  - Case Frames – *Fillmore* – FrameNet
  - Lexical Conceptual Structure – *Jackendoff* – LCS
  - Proto-Roles – *Dowty* – PropBank
  - English verb classes (diathesis alternations) - *Levin* - VerbNet
- Manual Semantic Annotation
- Automatic Semantic annotation
- Parallel PropBanks and Event Relations

Ask Jeeves – filtering w/ POS tag

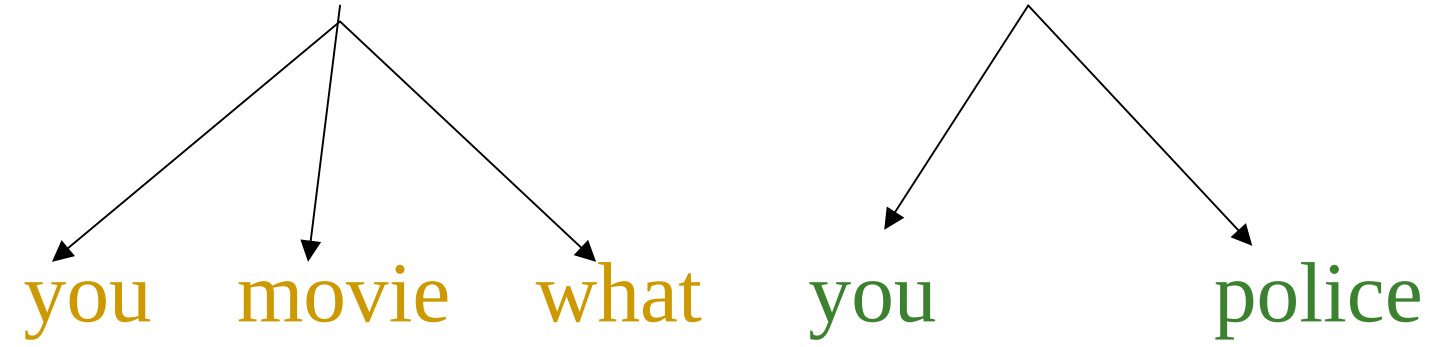
What do you call a successful movie?

- Tips on Being a **Successful Movie** Vampire ... I shall **call** the police.
- Successful Casting Call & Shoot for "Clash of Empires" ... thank everyone for their participation in the making of yesterday's movie.
- Demme's casting is also highly entertaining, although I wouldn't go so far as to **call** it **successful.** This **movie's** resemblance to its predecessor is pretty vague...
- VHS Movies: Successful Cold Call Selling: Over 100 New Ideas, Scripts, and Examples from the Nation's Foremost Sales Trainer.

Filtering out “call the police”

Different senses,
- different syntax,
- different kinds of participants,
- different types of propositions.

call(you,movie,what) ≠ call(you,police)



English lexical resource is required

AskJeeves: Who do you call for a good electronic lexical database for English?

WordNet – Princeton

(Miller 1985, Fellbaum 1998)

On-line lexical reference (dictionary)

- Nouns, verbs, adjectives, and adverbs grouped into synonym sets
- Other relations include hypernyms (ISA), antonyms, meronyms
- Typical top nodes - 5 out of 25
  - *(act, action, activity)*
  - *(animal, fauna)*
  - *(artifact)*
  - *(attribute, property)*
  - *(body, corpus)*

WordNet – Princeton

(Miller 1985, Fellbaum 1998)

- Limitations as a computational lexicon
  - Contains little syntactic information
  - No explicit lists of participants
  - Sense distinctions very fine-grained,
  - Definitions often vague
- Causes problems with creating training data for supervised Machine Learning – SENSEVAL2
  - Verbs > 16 senses (including *call*)
  - Inter-annotator Agreement ITA 71%,
  - Automatic Word Sense Disambiguation, WSD 64%

WordNet – call, 28 senses

1. name, **call** -- (assign a specified, proper name to;
   "They named their son David"; …)
   -> LABEL
2. **call**, telephone, call up, phone, ring -- (get or try to get into communication (with someone) by telephone;
   "I tried to call you all night"; …)
   ->TELECOMMUNICATE
3. **call** -- (ascribe a quality to or give a name of a common noun that reflects a quality;
   "He called me a bastard"; …)
   -> LABEL
4. **call**, send for -- (order, request, or command to come;
   "She was called into the director's office"; "Call the police!")
   -> ORDER

WordNet: - call, 28 senses, groups

WN5, WN16,WN12

Loud cry

WN15 WN26

Bird or animal cry

WN3 WN19

WN1 WN22

Label

WN4 WN 7 WN8 WN9

Request

WN20 WN25

Call a loan/bond

WN18 WN27

Challenge

WN6 WN23

Visit

WN2 WN 13

WN28

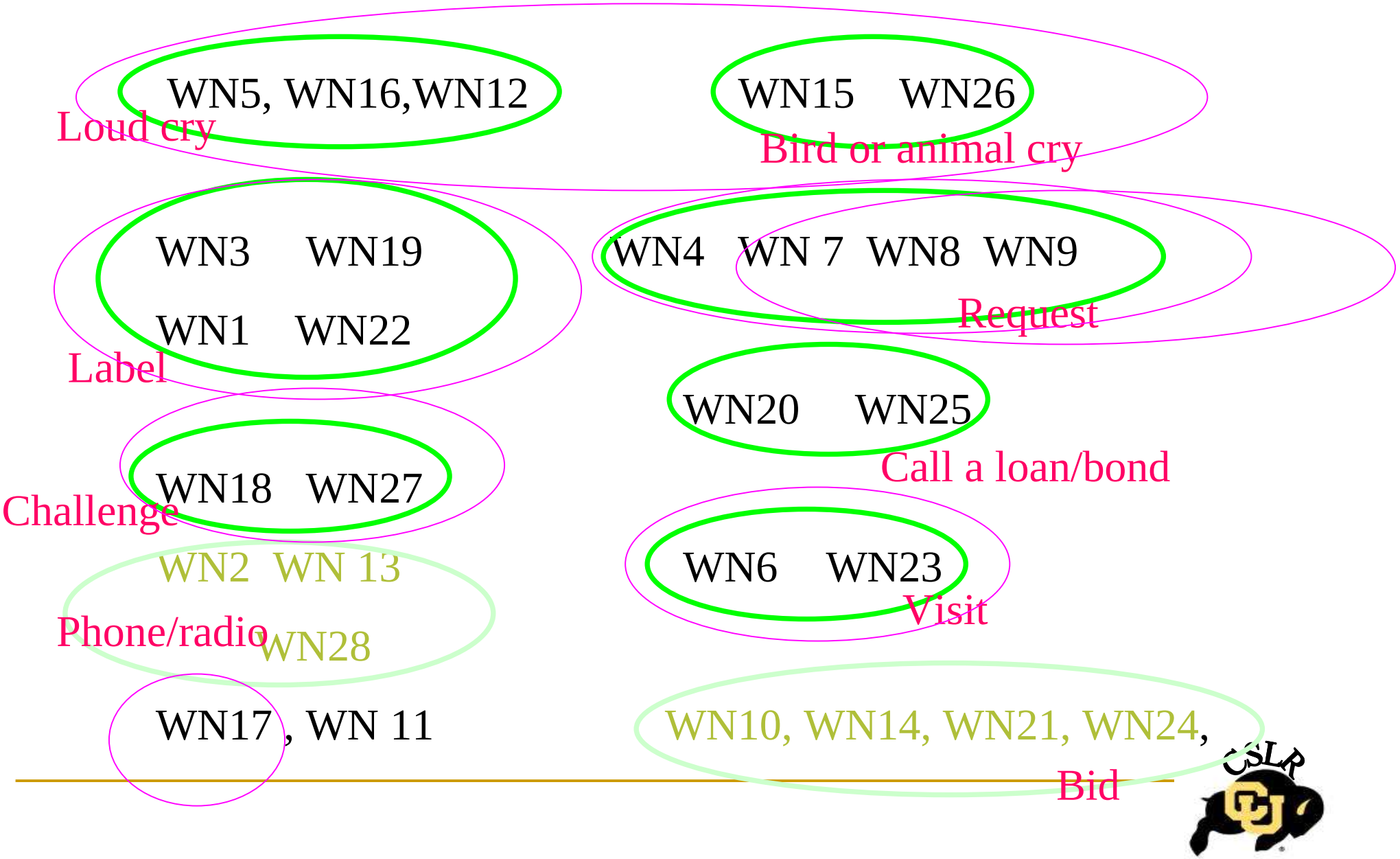
Phone/radio

WN17 , WN 11

WN10, WN14, WN21, WN24,

Bid

WordNet: - call, 28 senses, groups



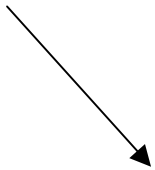
Overlap between Groups and Framesets – 95%

Palmer, Dang & Fellbaum, NLE 2004

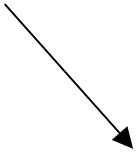
Sense Hierarchy

(Palmer, et al, SNLU04 - NAACL04, NLE06, Chen, et. al, NAACL06)

- PropBank Framesets – ITA >90%
  coarse grained distinctions
  20 Senseval2 verbs w/ > 1 Frameset
  Maxent WSD system, 73.5% baseline, 90%



  - Sense Groups (Senseval-2) - ITA 82%
    Intermediate level
    (includes Levin classes) – 69%



    - WordNet – ITA 73%
      fine grained distinctions, 64%

Tagging w/groups, ITA 90%, 200@hr, Taggers - 86.9%

Criteria to split Framesets

- Semantic classes of arguments, such as animacy vs. inanimacy

Serve 01. Act, work

- Group 1: function (*His freedom served him well*)
- Group 2: work (*He served in Congress*)

Criteria to split Framesets

- Syntactic variation of arguments

See 01. View

- Group 1: Perceive by sight

  (*Can you see the bird?*)

- Group 5: determine, check

  (*See whether it works*)

Criteria to split Framesets

- Optional Arguments

leave 01. Move away from

- Group 1: depart (*Ship leaves at midnight*)
- Group 2: leave behind (*She left a mess*.)

An example of sense mapping: ‘serve’

Frameset id = ***serve.01***	Sense Groups
serve 01: *Act, work* Roles: Arg0: worker Arg1: job, project Arg2: employer	GROUP 1: WN1 (function,‘The tree stump serves as a table’) WN3 (contribute to,‘the scandal served to increase ..) WN12 (answer, ‘Nothing else will serve’)
	GROUP 2: WN2 (do duty, ‘She served in Congress’) WN13 (do military service)
	GROUP 3: WN4 (be used,‘the garage served to shelter horses’) WN8 (promote, ‘their interests are served’) WN14 (service, mate with)
	GROUP 5: WN7 (devote one’s efforts. ‘serve the country’) WN10 (attend to, ‘May I serve you?’)

Goals – Ex. Answering Questions

- *Similar concepts*
  - *Where are the grape arbors located?*
  - *Every path from back door to yard was covered by a grape-arbor, and every yard had fruit trees.*

WordNet – cover, 26 senses

1. **cover** -- (provide with a covering or cause to be covered; "*cover the grave with flowers*") -> ??

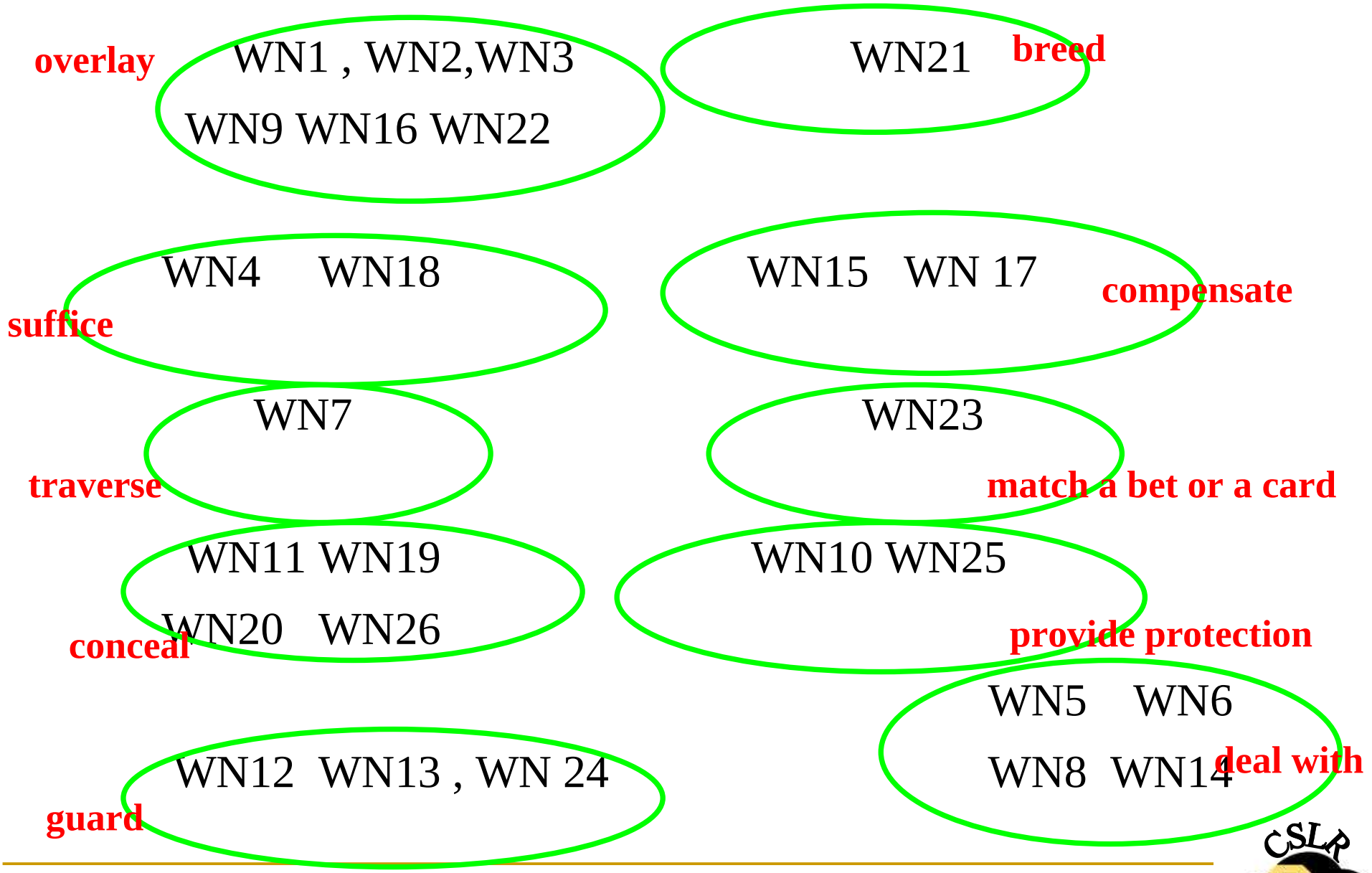
2. **cover**, spread over -- (form a cover over; *"The grass covered the grave")* ->TOUCH

4. **cover** -- (provide for; *"The grant doesn't cover my salary")* -> SATISFY, FULFILL

7. traverse, track, **cover**, cross, pass over, get over, get across, cut through, cut across -- ("The caravan covered almost 100 miles each day") -> TRAVEL

8. report, **cover** -- (be responsible for reporting the details of, as in journalism; "The cub reporter covered New York City") -> INFORM

WordNet: - cover, sense grouping



Frame File example: *cover.01* –

PropBank instances mapped to VerbNet

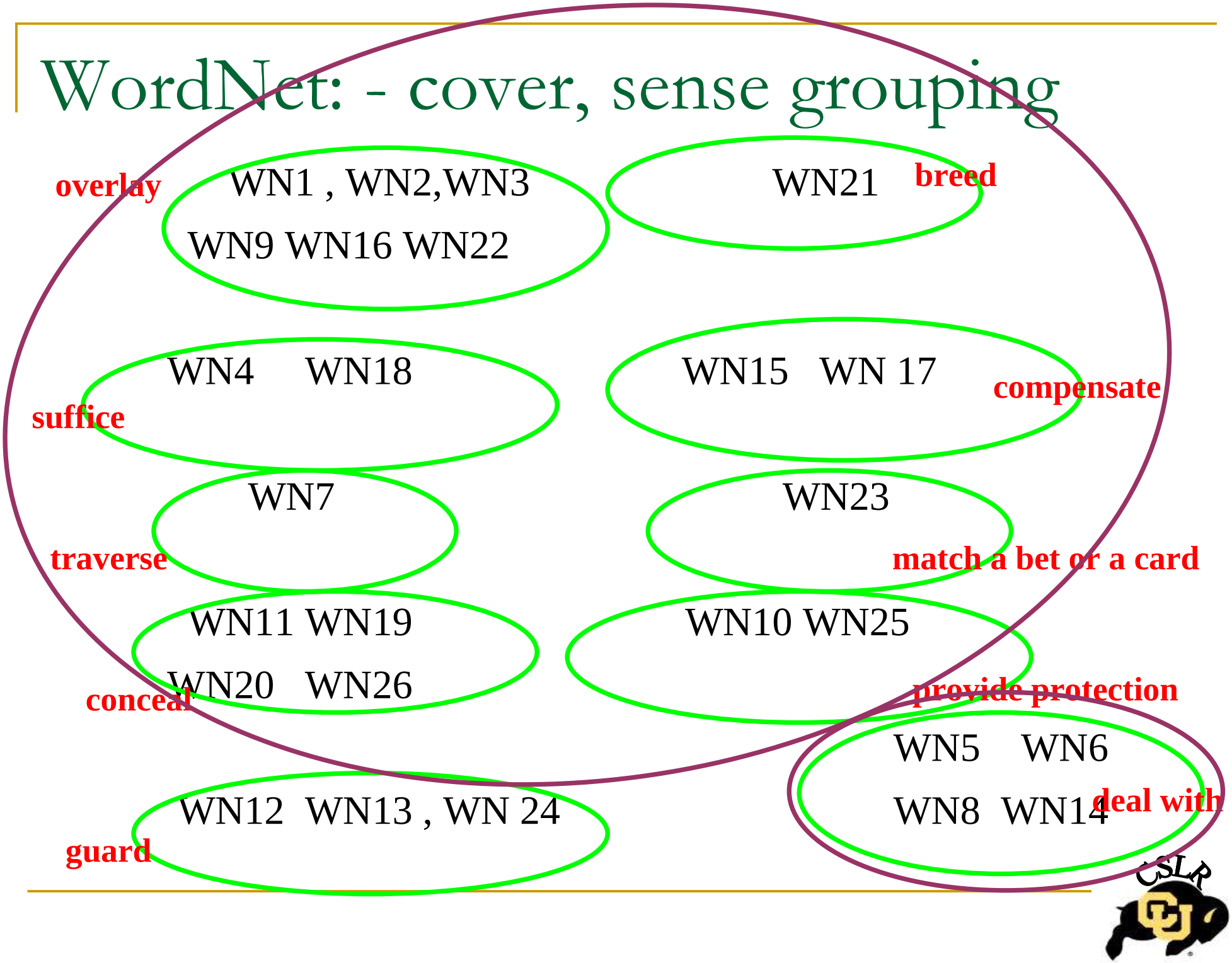
Roles:

Arg0: coverer
Arg1: thing covered
Arg2: cover

Example: She covered her sleeping baby with a blanket.

Arg0:	Agent	*She*
REL:		*covered*
Arg1:	Destination	*her sleeping baby*
Arg2:	Theme	*with a blanket*

WordNet: - cover, sense grouping



VerbNet - *cover contiguous_location-47.8*

- **WordNet Senses:**
  border(1,2,5),…,cover(2), edge(3),…,

- **Thematic Roles:** Theme [+concrete],
  Theme [+concrete]

- **Frames with Semantic Roles**

  "Italy borders France"

  Theme1 V Theme2

  contact(during(E),Theme1,Theme2)
  exist(during(E),Theme1) exist(during(E),Theme2)

VerbNet – *cover fill-9.8*

- **WordNet Senses: …,** cover(1,2, 22, 26),…, staff(1),

- **Thematic Roles:** Agent [+animate]
  Theme [+concrete],
  Destination [+location, +region]

- **Frames with Semantic Roles**

  "The employees staffed the store"

  " The grape arbors covered every path"

  Theme V Destination

  location(E,Theme,Destination)
  location(E,grape_arbor,path)

Goals – Lexical chaining for Q/A

- *Similar concepts*
  - *Where are the grape arbors located?*
  - *Every path from back door to yard was covered by a grape-arbor, and every yard had fruit trees.*

    No lexical overlap w/ WordNet 2.0 entries

    4 senses for “locate” and 26 for “cover.”

  VerbNet gives us two classes for *cover*, one with contact and one with location.

  Which one?

FrameNet: Telling.*inform*

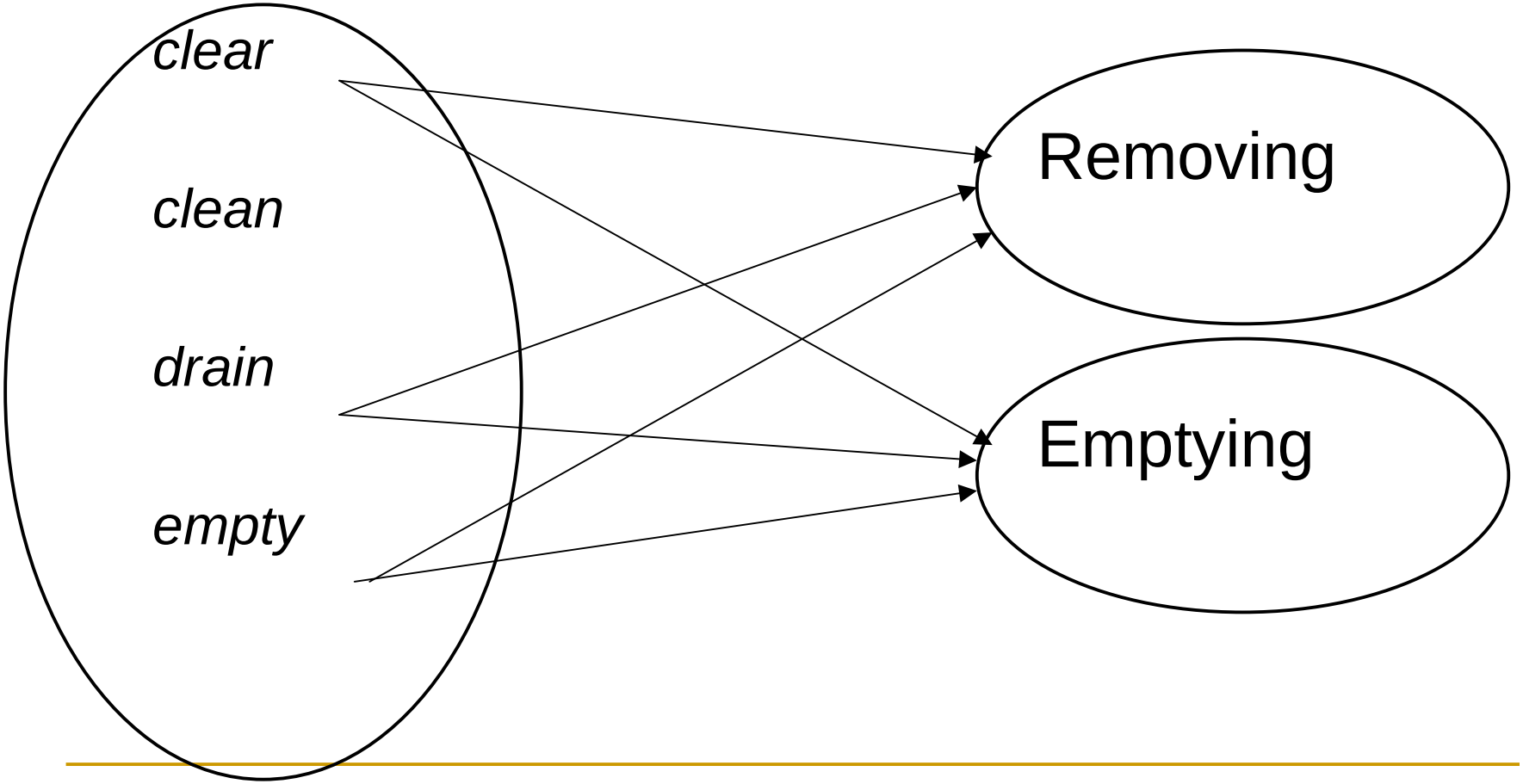
Time	In 2002,
Speaker	the U.S. State Department
Target	INFORMED
Addressee	North Korea
Message	that the U.S. was aware of this program , and regards it as a violation of Pyongyang's nonproliferation commitments

FrameNet/PropBank:Telling.*inform*

Time	ArgM-TMP	In 2002,
Speaker –	Arg0 *(Informer)*	the U.S. State Department
Target –	REL	INFORMED
Addressee –	Arg1 *(informed)*	North Korea
Message –	Arg2 *(information)*	that the U.S. was aware of this program , and regards it as a violation of Pyongyang's nonproliferation commitments

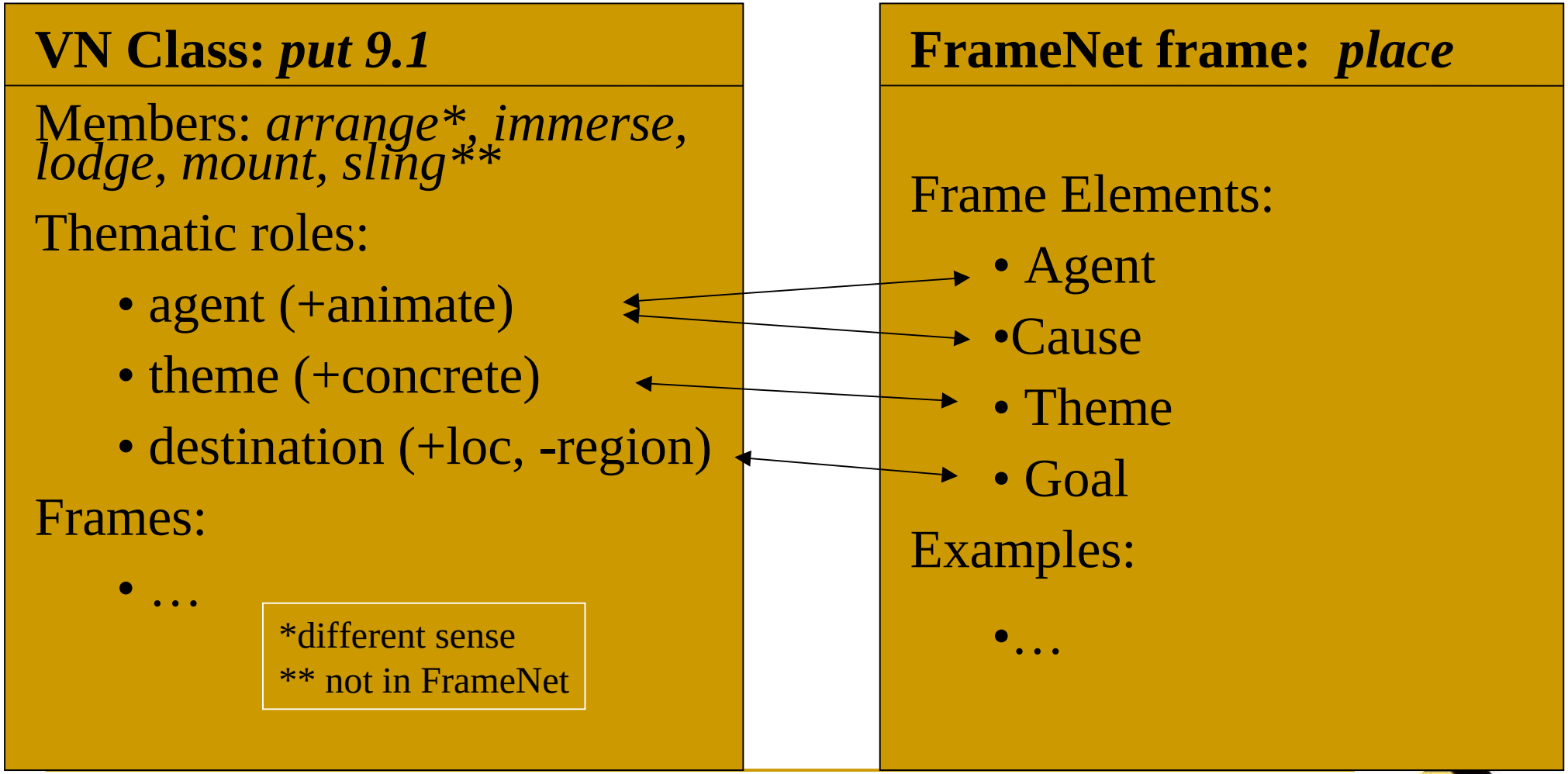
Mapping Issues (2)
VerbNet verbs mapped to FrameNet

- *VerbNet clear-10.3*
- FrameNet Classes



Mapping Issues (3)

VerbNet verbs mapped to FrameNet



FrameNet frames for Cover

- *"overlay"* Filling (also Adorn, Abounding with)

Theme fills Goal/Location by means of Agent or Cause. *[She* $_{Agent}$*] covered [her sleeping child* $_{Goal}$*] with [a blanket* $_{Theme}$*].*

- "deal with" Topic - Text or Discourse that a Communicator produces about a Topic

[Local news $_{Communicator}$*] will cover these [events* $_{Topic}$*]*

- " hide" Eclipse - An Obstruction blocks an Eclipsed entity from view, *[This make-up* $_{Obstruction}$*] will cover [your acne* $_{Eclipsed\ Entity}$*]*

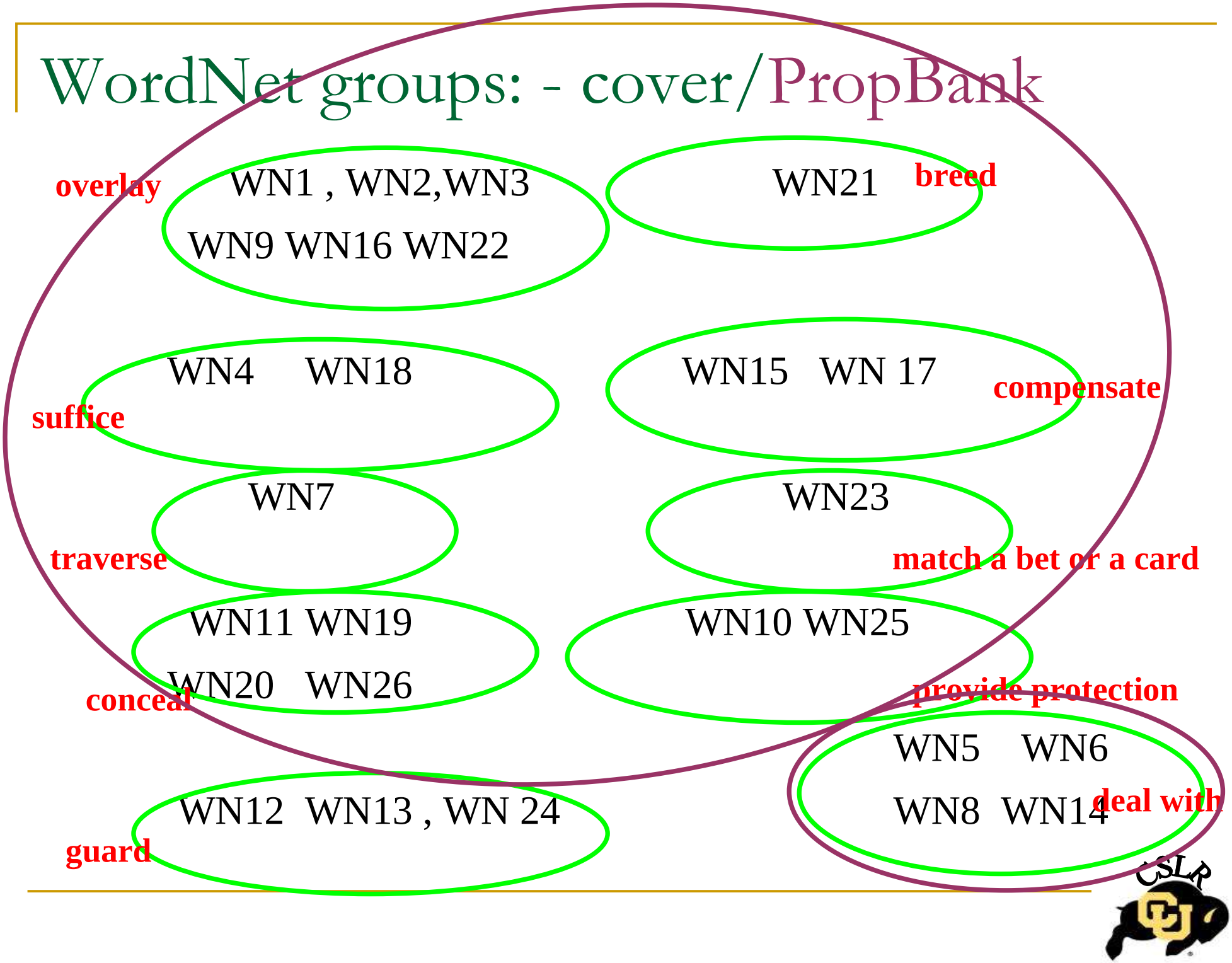
Mapping Resources

- PropBank/VerbNet/FrameNet
- PropBank/WordNet sense groupings

- How well do sense groupings, VerbNet classes, and FrameNet frame memberships overlap?

WordNetGroups: - cover/VerbNet

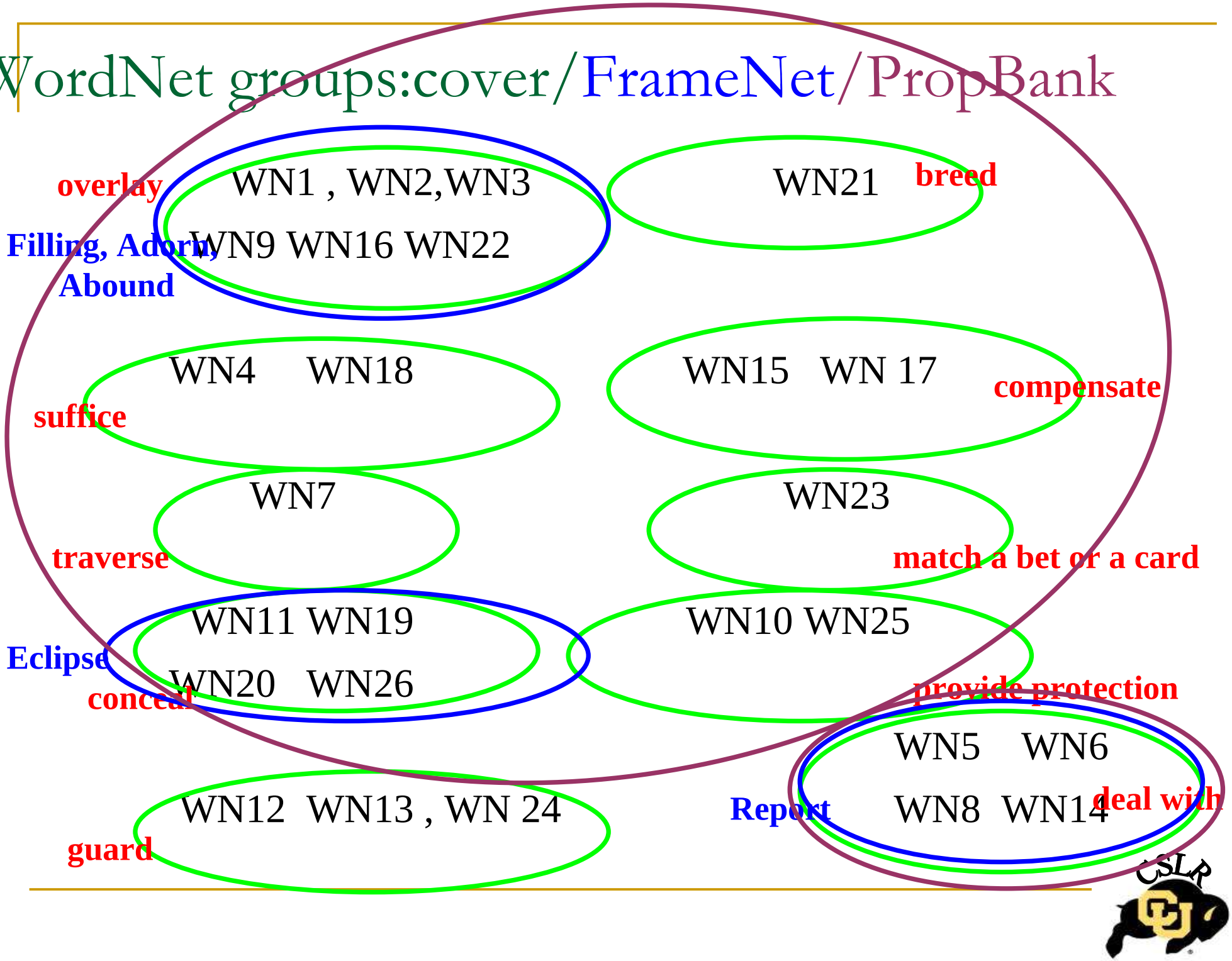
WordNet groups: - cover/PropBank



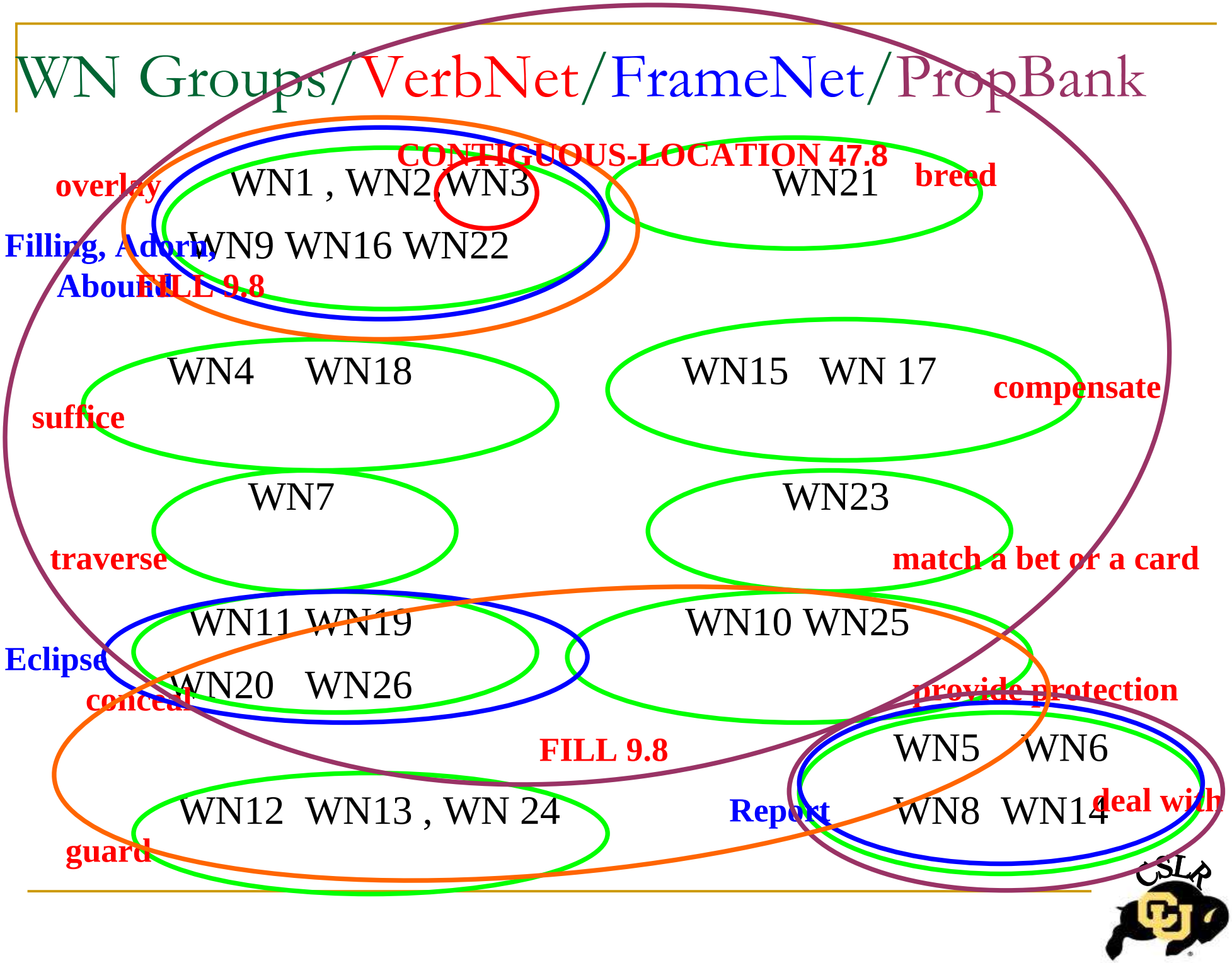
WordNet groups:cover/VerbNet/PropBank

WordNet groups: - cover/FrameNet

WordNet groups:cover/FrameNet/PropBank



WN Groups/VerbNet/FrameNet/PropBank



How far have we come?

- We now have predicate argument structures with senses and ontology links, but no relations between them
- We need to identify both verbal and nominal events so that we can define relations between them – co-referential, temporal and discourse relations.
- This will also simplify mapping between a verbal expression in one language and a nominal expression in another.

Outline

- Linguistic Theories of semantic representation
  - Case Frames – *Fillmore* – FrameNet
  - Lexical Conceptual Structure – *Jackendoff* – LCS
  - Proto-Roles – *Dowty* – PropBank
  - English verb classes (diathesis alternations) - *Levin* - VerbNet
- Manual Semantic Annotation
- Automatic Semantic annotation
- Parallel PropBanks and Event Relations

A Parallel Chinese-English PropBank II

Martha Palmer, Nianwen Xue, Olga Babko-Malaya, Jinying Chen,

University of Pennsylvania

&

University of Colorado

Proposition Bank I: An Example

Mr. Bush met him privately, in White House, on Thursday.

Rel: met

Arg0: Mr. Bush

Arg1: him

ArgM-MNR: privately

ArgM-LOC: in White House

ArgM-TMP: on Thursday

- What other layers of annotation do we need to map sentences into propositions?

∃e meeting(e) & Arg0(e, Mr.Bush) & Arg1(e, he) & MNR(e, privately) & LOC(e, 'in White House') & TIME(e, 'on Thursday')

PropBank II – English/Chinese (100K)

We still need relations between events and entities:

- Event ID's with event coreference
- Selective sense tagging
  - Tagging nominalizations w/ WordNet sense
  - Grouped WN senses - selected verbs and nouns
- Nominal Coreference
  - not names
- Clausal Discourse connectives – selected subset

Level of representation that reconciles many surface differences between the languages

Criteria for grouping WN senses: relation to events

'development':

- Group 1 (Event) The act of growing, evolving, building, improvement (WordNet senses: 1, 2, 4, 7, 8)

  "The development of the plan took only ten years". "The development of an embryo is a complicated process". "If development of your pictures takes more that one hour - it's free".

- Group 2. The End Product. The result of growing, evolving, building, improvement (WordNet Sense 5)

  "That housing development is beautiful".

Eventuality Variables

- Identify eventualities
  - Aspectual verbs do not introduce eventualities

    New loans continued to slow.
  - Some nominals do introduce eventualities

    The real estate development ran way over budget.

Aspectual Verbs

New loans continued to slow.

PB annotation:

rel: continue
Arg1: [New loans] [to slow]

rel: slow
Arg1: New loans

PB annotation with events:

m1 - rel: continue
Arg1: e2

e2 – rel: slow
Arg1: New loans

Identifying Eventuality Arguments

Society for Savings Bancorp saw its stock rise.

Event id	Rel	Arg0	Arg1
e23	rise		Its stock
e16	see	Society for Savings Bancorp	e23

Annotation on selected classes of verbs:

- **aspectual verbs**
- **verbs of perception**
- **verbs like ‘happen’, ‘occur’, ‘cause’**
- **selected using VerbNet**

-PTB: 16093 instances; ECTB: 1346 instances

Eventuality coreference

A successor was n't named [*-1] , which [*T*-35] fueled speculation that Mr. Bernstein may have clashed with S.I. Newhouse Jr.

Event id	Rel	Arg1	Arg2
e23	named	a sucessor	
e16	fueled	[*T*-35] -> which -> e23	speculation that Mr. Bernstein may have clashed with S.I. Newhouse Jr

Nominal Coreference

- Restricted to direct coreference, or identity relation

- Pronominal coreference
- Definite NPs (including temporals), but only identity relations.

John spent [three years] in jail. In [that time]...

**Morril Hall does not have [a bathroom] or [it]'s in a funny place*

Classification of pronouns

- 'referring'
  [John Smith] arrived yesterday. [He] said that...

- ‘bound'
  [Many companies] raised [their] payouts by more than 10%

- ‘event‘
  Slowing [e] the economy is supported by some Fed officials, [it] is repudiated by others.

- ‘generic'
  I like [books]. [They] make me smile.

Annotation of free traces

- Free traces – traces which are not linked to an antecedent in PropBank

- Arbitrary

  Legislation to lift the debt ceiling is ensnarled in the fight over [*]–ARB cutting capital-gains taxes

- Event

  The department proposed requiring (e4) stronger roofs for light trucks and minivans , [*]-e4 beginning with 1992 models

- Imperative

  All right, [*]-IMP shoot.

- **1K instances of free traces in a 100K corpus**

Parallel Chinese/English PropBank II

- The English annotation is all done on the PTB and the English side of the 100K parallel C/E corpus
- Chinese PB II annotation projects
  - Sense group tagging
  - Event identification and event coreference
  - Discourse connectives

Event IDs – Parallel Prop II (1)

- Aspectual verbs do not receive event ID's:
  - 今年/this year 中国/China 继续/continue 发挥/play 其/it 在/at 支持/support 外商/foreign business 投资/investment 企业/enterprise 方面/aspect 的/DE 主/main 渠道/channel 作用/role

"This year, the Bank of China will continue to play the main role in supporting foreign-invested businesses."

Event IDs – Parallel Prop II (2)

- Nominalized verbs do:
  - *He will probably be extradited to the US for trial.*
    done as part of sense-tagging
    (all 7 WN senses for "trial" are events.)

  - 随着/with 中国/China 经济/economy 的/DE 不断/continued 发展/development…

*"With the continued **development** of China's economy…"*

The same events may be described by verbs in English and nouns in Chinese, or vice versa. Event ID's help to abstract away from POS tag

Event reference – Parallel Prop II

- Pronouns (overt or covert) that refer to events:

[This] is gonna be a word of mouth kind of thing.

这些/these 成果/achivements 被/BEI 企业/enterprise 用/apply (e15) 到/to 生产/production 上/on 点石成金/spin gold from straw， *pro*-e15 大大/greatly 提高/improve 了/le 中国/China 镍/nickel 工业/industry 的/DE 生产/production 水平/level 。

*"These achievements have been applied **(e15)** to production by enterprises to spin gold from straw, **which-e15** greatly improved the production level of China's nickel industry."*

Prerequisites:

- pronoun classification
- free trace annotation

Chinese PB II: Sense tagging

- Much lower polysemy than English
  - Avg of 3.5 (Chinese) vs. 16.7 (English)

    Dang, Chia, Chiou, Palmer, COLING-02
  - More than 2 Framesets

    62/4865 (250K) Ch vs. 294/3635 (1M) English

- Mapping Grouped English senses to Chinese

  (English tagging - 93 verbs/168 nouns, 5000+ instances)
  - Selected 12 polysemous English words

    (7 verbs/5 nouns)
  - For 9 (6 verbs/3 nouns), grouped English senses map to unique Chinese translation sets (synonyms)

Mapping of Grouped Sense Tags to Chinese

raise – translations by group

Mapping of Grouped Sense Tags to Chinese

- **Zhejiang|浙江zhe4jiang1 will|将jiang1 raise|提高ti2gao1 the level|水平shui3ping2 of|的de opening up|开放kai1fang4 to|对dui4 the outside world|外wai4.** （浙江将提高对外开放的水平。）
- **I|我wo3 raised|仰yang3 my|我的wo3de head|头tou2 in expectation|期望qi1wang4.**（我仰头望去。）
- **…, raising|筹措chou2cuo4 funds|资金zi1jin1 of|的de 15 billion|150亿yi1ban3wu3shi2yi4 yuan|元yuan2 (…筹措资金150亿元。)**
- **The meeting|会议hui4yi4 passed|通过tong1guo4 the "decision regarding motions"|议案yi4an4 raised|提ti4 by 32 NPC|人大ren2da4 representatives|代表dai4biao3** （会议通过了32名人大代表所提的议案。）

Discourse connectives: The Penn Discourse TreeBank

- WSJ corpus (~1M words, ~2400 texts)
  http://www.cis.upenn.edu/~pdtb
  Miltsakaki, Prasad, Joshi and Webber, LREC-04, NAACL-04 Frontiers
  Prasad, Miltsakaki, Joshi and Webber ACL-04 Discourse Annotation

- Chinese: 10 explicit discourse connectives that include subordination conjunctions, coordinate conjunctions, and discourse adverbials.

- Argument determination, sense disambiguation

[*arg1* 学校/school 不/not 教/teach 理财/finance management], [*conn* 结果**/as a result**] [*arg2* 报章/newspaper 上/on 的/DE 各/all 种/kind 专栏/column 就/then 成为/become 信息/information 的/DE 主要/main 来源/source]。

"*The school does not teach finance management. **As a result**, the different kinds of columns become the main source of information.*"

Summary of English PropBanks

Olga Babko-Malaya, Ben Snyder

Genre	*Words*	*Frames Files*	*Frameset Tags*	*Released*	*Prop2*
Wall Street Journal* (Penn TreeBank II)	1000K	< 4000	700+	March, 04	
English Translation of Chinese TreeBank *	100K	<1500		Dec, 04	Aug, 05
Xinhua News DOD funding	250K	< 6000	200	Dec, 04	Dec, 05 (100K)
Sinorama NSF-ITR funding	150K	< 4000		July, 05	
Sinorama, English corpus NSF-ITR funding	250K	<2000		Dec, 06	

*DOD funding

NSF Grant – Unified Linguistic Annotation

- James Pustejovsky, PI,
- Co-PI's - Martha Palmer, Adam Meyers, Mitch Marcus, Aravind Joshi, Jan Weibe
- Unifying Treebank, PropBank, NomBank, Discourse Treebank, Opinion Corpus, Coreference
- Events with relations between them!

Goal

- Next step – Inferencing
- Prerequisites
  - Real propositions, not just predicate argument structures
  - Links to an ontology

Event relations - Example

- *The White House said President Bush has approved duty-free treatment for imports of certain types of watches that aren't produced in "significant quantities" in the U.S., the Virgin Islands and other U.S. possessions. The action came in response to a petition filed by Timex Inc. for changes in the U.S. Generalized System of Preferences. Previously, watch imports were denied such duty-free treatment.*