

Workshop: Urdu WordNet

Farhat Abdullah
Ayesha Zafar
Afia Mahmood



Centre for Language Engineering
Al-Khwarizmi Institute of Computer Science, University of
Engineering and Technology Lahore, Pakistan

Problems of Translation

“Elephant was lifting a stone with its **trunk**”

باتھی اپنی **سونڈ** سے پتھر اٹھا رہا تھا



trunk= دھڑ، صندوق، ڈگی، تنا، سونڈ

Problems of Translation

- Finding the right word in the target language

--the sense of a word that is intended by the
writer of the source text

--the appropriate word-meaning mapping in the
target text

Webster

<http://www.merriam-webster.com/dictionary/trunk>

1. The main stem of a tree
2. The Human or animal body
3. Central part of anything
4. Large rigid piece of luggage
5. A superstructure over a ship
6. The long muscular proboscis of the elephant

Cambridge Dictionary Online

http://dictionary.cambridge.org/dictionary/british/trunk_1?q=trunk

1. The thick main stem of a tree, from which its branches grow
2. The main part of a person's body, not including the head, legs or arms

Oxford Dictionary

<http://oxforddictionaries.com/definition/english/trunk?q=trunk>

1. The main woody stem of a tree
2. The main part of an artery, nerve, or other anatomical structure
3. A person's or animal's body apart from the limbs and head
4. The elongated, prehensile nose of an elephant
5. A large box with a hinged lid for storing or transporting clothes and other articles
6. The boot of a car

Limitation of Dictionaries

- Compiled (alphabetically) on historical (diachronic) principles
- Order of entries is not the same
- Tag/ code number of senses is not the same
- The number of senses are different per category in different dictionaries

Need

- An aid to search lexicons conceptually, rather than alphabetically
- Entries are organized in a definite order
- Specific tag/code number is assigned to a sense
- Pre-defined number of senses for each category

Purpose of Development

- Globalization requires more texts and speech to be translated faster across more languages
- Machine translation is difficult , expensive and time-consuming
- Machine translation is of low quality. Often unacceptable

WordNet

- Lexical database
- Grouped into sets of cognitive synonyms
- each expressing a distinct concept (synsets)
 - Nouns, verbs, adjectives and adverbs
- Useful tool for linguistics and natural language processing

Components of WordNet

- **Synsets:** It is set of different words having same semantic concept
 - exchange of any of these words does not change the semantic property of an sentence

{چادر، میز پوش}

{چادر، دوپٹہ، شال، امبر، ردا}

{trunk, tree trunk, bole}

{trunk, torso , body}

{trunk, luggage compartment, automobile trunk}

{trunk, proboscis}

Components of WordNet (contd.)

Unique ID: Every sense has a unique ID which is assigned to it after mapping the accurate sense

Category: Clearly defined and managed systematically

Concept: An explained and comprehensive statement is given to elaborate the semantic value of the sense

Example: Any word from the synset is used in an example to further elaborate the sense

WordNet DB {Synsets, Unique ID, Category, Concept, Example}

1. {12995758} <noun.plant> **trunk#1, tree trunk#1, bole#2** -- (the main stem of a tree; usually covered with bark; the bole is usually the part that is commercially useful for lumber)
 2. {04438323} <noun.artifact> **trunk#2** -- (luggage consisting of a large strong case used when traveling or for storage)
 3. {05480848} <noun.body> **torso#1, trunk#3, body1#4** -- (the body excluding the head and neck and limbs; "they moved their arms and legs and bodies")
 4. {03655285} <noun.artifact> **luggage compartment#1, automobile trunk#1, trunk#4** -- (compartment in an automobile that carries luggage or shopping or tools; "he put his golf bag in the trunk")
- **5. {02430617} <noun.animal> **proboscis#2, trunk#5** -- (a long flexible snout as of an elephant)

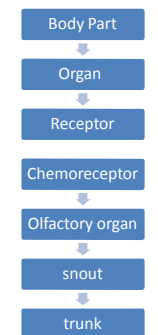
Some relations in WordNet

• Lexical relations

- Synonymy
- Antonymy

• Semantic Relations

- **hyponymy, hyponymy**
- or **ISA relation**



Uses of WordNet

- Word sense disambiguation
- Information retrieval
- Automatic text classification
- Automatic text summarization
- Machine translation
- Automatic crossword puzzle generation
- Determine the semantic similarity between words

WordNet: History

- 1985: a group of psychologists and linguists start to develop a “lexical database”
- Princeton University
- Theoretical basis: results from
- Psycholinguistics and psycholexicology
- What are properties of the “mental lexicon”?

Princeton WordNet

- In the absence of an easily available electronic dictionary
- An extensive electronic dictionary of the English language
- Comprising more than 200,000 word-meaning-pairs
- Various off springs mapping WordNet's achievements onto languages other than English

Versions of Princeton WordNet

- 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.7.1, 2.0, 2.1,3.0
- **2.0, 2.1:** all nouns are in one tree under "entity" in "noun.Tops"
- WordNet URL is now "wordnet.princeton.edu"
- **2.1, 3.0:** some changes were made to the graphical interface and WordNet library with regard to adjective and adverb searches
- A separate "Related Noun" search was inserted for adjectives

<http://wordnet.princeton.edu/wordnet/download/old-versions/>

WordNets for Other Languages

- Idea has been widely adapted
- by "translating" Princeton WordNet
 - Lexical relations in general are universal
- **Euro WordNet:** English, Dutch, German, French, Spanish, Italian, Czech, Estonian
- **BalkaNet:** Romanian, Bulgarian, Turkish, Slovenian, Greek, Serbian
- **Indo WordNet:** is a linked lexical knowledge base of WordNets of 18 scheduled languages of India, viz.^[27]

Global WordNet

- A free, public and non-commercial organization
- It provides a platform for discussing, sharing and connecting WordNets for all languages in the world.
- It promotes the standardization of WordNet across different languages
- To ensure its uniformity in enumerating the different synsets in human languages

Approaches to Develop WordNet

- **Expand approach:** translates WordNet synsets to another language and take over the structure
 - easier and more efficient method
 - compatible structure with WordNet
 - vocabulary and structure is close to WordNet but also biased
 - can exploit many resources linked to WordNet
- **Merge approach:** creates an independent WordNet in another language and align it with WordNet by generating the appropriate translations
 - more complex and labor intensive
 - different structure from WordNet
 - language specific patterns can be maintained, i.e. very precise substitution patterns

Urdu WordNet

- The purpose of the development of Urdu WordNet is to provide a lexical resource for Urdu language that can be used in natural language processing
- The WordNet is being developed specifically to align with local linguistic, cultural, religious and other contexts
- To build Urdu language WordNet merge approach has been used

Practice Session

Step 1: Category

- Determine the Part of Speech (POS) tags of the word with the help of Urdu Dictionary

<http://www.clepk.org/oud/>

کھانا



کھانا [کھا + نا] (پراکرت)

پراکرت زبان سے ماخوذ ہے۔ اردو میں عربی رسم الخط کے ساتھ بطور اسم اور فعل متعدی استعمال ہوتا ہے۔ 1609ء کو "قطب مشنری" میں ضمیمہ میں مستعمل ملتا ہے۔

اسم نکرہ (مذکر - واحد)

واحد غیر ندانی: کھائے [کھا + نے]

جمع: کھائے [کھا + نے]

جمع غیر ندانی: کھاتوں [کھا + توں (واؤ مجہول)]

1. طعام، غذا، خوراک۔

"ایک دن بابائے اردو کو معلوم ہوا کہ علی شیر حتمی کو بخار ہو گیا ہے تو ایسے بے چین ہوئے کہ دن میں تین بار دیکھنے لگے اور رات کے وقت کھانا بھی نہیں کھایا۔" (1990ء، اردو نامہ، لاہور، مارچ، 21)

2. [دکن] خشکا، (شکر) ضیافت، دعوت۔

(ماخوذ: اصطلاحات پیشہ وران، 163:3، فرہنگ آصفیہ، نور اللغات)

انگریزی ترجمہ

eating, food, meal, dinner

فعل متعدی

1. کسی شے کو جبا کر حلق سے نیچے اُتارنا، نگلنا۔
"مُو تو شام کو اپنے ہاتھ سے کھانا بھی نہیں کھاتا۔" (1982ء، پرایا گہر، 103)

2. [مجازاً] غم سینا، مصیبت اٹھانا۔
آشک پیتا ہوں تو کھانا ہوں غم میں اٹھ پیر
اور اوقات بسر ہوتے ہی صورت کیا ہے (1932ء، بے نظیر، کلام بے نظیر، 167)

3. [مجازاً] گنجائش رکھنا، سمائی ہونا، کھپ جانا۔
"اس کے گرد چھوٹی توپیں جو تیز ہ ہونڈ کا گولہ کھاتی تھیں۔" (1889ء، حرم سرا، 134:1)

4. تلوار، لکڑی یا کسی اور چیز کی ضرب کھانا، زخم کھانا، چوٹ یا مار کھانا۔
زخم جس سے تیری تلوار کا کھایا نہ گیا
سر کبھی اس سے شیبوں میں اٹھایا نہ گیا (1870ء، دیوان امیر، 43:3)

5. (نصیب یا مقدر کی گردش) سبنا یا برداشت کرنا، ٹھوکرین کھانا۔
ڑاہ پر خود آجائے آخر ٹھوکرین کچھ دن کھائے دو
دیوانہ ہے دل سے نہ الجھو آؤ بھی ہو گا جاتے دو (1926ء، دیوان سنی، 36)

6. [مجازاً] تباہ بریاد کرنا، مٹانا، معلوم کر دینا۔
تجب سے چمن چھتا ہے یہ حال ہو گیا ہے
دل غم کو کھا رہا ہے، غم دل کو کھا رہا ہے (1924ء، بانگ درا، 24)

7. استعمال میں لانا، صرف کر ڈالنا۔
"حضور ہمارے صاحب کے جوتے اس کلاس چمڑے کے ہیں، زیادہ پالش کھاتے ہیں۔" (1947ء، مسلمانوں فرحت، 43:3)

8. بڑب کر جانا، نگل جانا، قبضہ کر لینا، بضم کر لینا۔
"اب سنے میں آ رہا ہے کہ ناگپور کے ہونسل خانان کی مرتبہ ریاست کو بھی کھانا چاہتا ہے۔" (1946ء، جہانسی کی رانی، 21)

9. ڈسنا، مٹانا، تکلیف دینا۔
"جو شخص اس لیے لوگوں سے بھیک مانگے اور اس ذریعہ سے اپنے مال کو بڑھائے تو یہ سوال قیامت کے دن اس کے چہرے پر ایک زخم ہوگا اور دوزخ میں گرم پتھر بن کر اس کو کھائے گا۔" (1965ء، مشکوٰۃ شریف، 434:1)

10. جی یا جان کے ساتھ پریشان کرنا، تنگ کرنا۔
پس میں کہتا ہوں اپنے گہر جاؤ
حضرت شوق تم نہ جی کھاؤ (1798ء، دیوان سوزِ قدیم، 462)

11. (قسم یا حلف) اٹھانا، عید کرنا۔
ٹازہ غم کھایا کئے ہم وہ ہیں پاکیزہ مزاج
اور تم کھاتے رہے چھوٹی قسم کھائی ہوئی (1884ء، آفتاب داغ، 142)

12. مطابقت رکھنا، ملنا۔
"یہ تمام عناصر مسلسل اور علی الترتیب مرتب نہیں ہوں گے بلکہ مجموعوں میں بٹ جائیں گے اور ان میں بعض دوسروں سے موافقت کھاتے معلوم ہوں گے۔" (1910ء، ادیب، نوامیس، 208)

13. رشوت لینا، اثر قبول کرنا (سردی یا اوس وغیرہ کا)، چائنا، کھوکھلا کر دینا، اٹھنا، صرف ہونا، سننا (گالی۔ (نور اللغات)

Step 1: Category

| Urdu ID | English ID | English Word | Category | Concept | Example | Synsets |
|---------|------------|--------------|----------|---------|---------|---------|
| 1 | | | N | | | کھانا |
| 2 | | | V | | | کھانا |

Step 1: Exercise

قیامت

| Urdu ID | English ID | English Word | Category | Concept | Example | Synsets |
|---------|------------|--------------|----------|---------|---------|---------|
| 1 | | | N | | | قیامت |
| 2 | | | Adj. | | | قیامت |

Step 2

- Select a sense to record for WordNet from Urdu Dictionary e.g.



کھانا [کھا + نا] (پراکرت)

پراکرت زبان سے ماخوذ ہے۔ اردو میں عربی رسم الخط کے ساتھ بطور اسم اور فعل متعدی استعمال ہوتا ہے۔ 1609ء کو "قطب مشنری" میں ضمیمہ میں مستعمل ملتا ہے۔

اسم نکرہ (مذکر - واحد)

واحد غیر ندانی: کھانے [کھا + نے]

جمع: کھانے [کھا + نے]

جمع غیر ندانی: کھانوں [کھا + نوں (واو مجہول)]

1. طعام، غذا، خوراک۔

"ایک دن بہانے اردو کو معلوم ہوا کہ علی شیر حتمی کو بخار ہو گیا ہے تو اسے بے چین ہونے کہ دن میں تین بار دیکھنے لگنے اور رات کے وقت کھانا بھی نہیں کھاتا۔" (1990ء، اردو نامہ، لاہور، مارچ، 21)

2. [دکن] خشک، (لشکری) ضیافت، دعوت۔

(ماخوذ: اصطلاحات پیشہ وران، 163:3، فریڈگ اصفیہ، نور اللغات)۔

انگریزی ترجمہ

eating, food, meal, dinner

فعل متعدی

1. کسی شے کو جبا کر حلق سے نیچے اُتارنا، نگلنا۔

"مُو تو شام کو اپنے ہاتھ سے کھانا بھی نہیں کھاتا۔" (1982ء، پرایا گہر، 103)

2. [مجازاً] غم سینا، مصیبت اٹھانا۔

آشک پینا ہوں تو کھانا ہوں غم میں اٹھ پھر

اور اوقات بسر ہونے کی صورت کیا ہے (1932ء، بے نظیر۔ کلام بے نظیر، 167)

3. [مجازاً] گنجائش رکھنا، سمائی ہونا، کھپ جانا۔

"اس کے گرد چھوٹی توپیں جو تیز ہ پونڈ کا گولہ کھاتی تھیں۔" (1889ء حرم سرا، 134:1)

4. تلوار، لکڑی یا کسی اور چیز کی ضرب کھانا، زخم کھانا، چوٹ یا مار کھانا۔

زخم جس سے تیری تلوار کا کھانا نہ گیا

سر کبھی اس سے شینوں میں اٹھایا نہ گیا (1870ء، دیوان امیر، 43:3)

5. (نصیب یا مقدر کی گردش) سبنا یا برداشت کرنا، ٹھوکرین کھانا۔

ڑاہ پر خود آجانے آخر ٹھوکرین کچھ دن کھانے دو

دیوانہ سے دل سے نہ الجھو آؤ بھی ہو گا جانے دو (1926ء، دیوان سنی، 36)

6. [مجازاً] تباہ بریاد کرنا، مٹانا، معلوم کر دینا۔

تجب سے چمن چھتا ہے یہ حال ہو گیا ہے

دل غم کو کھا رہا ہے، غم دل کو کھا رہا ہے (1924ء، بانگ درا، 24)

7. استعمال میں لانا، صرف کر ڈالنا۔

"حضور ہمارے صاحب کے جوتے اس کلاس چمڑے کے ہیں، زیادہ پالش کھاتے ہیں۔" (1947ء، مسلمانوں فرحت، 43:3)

Step 3: Concept

- Write the meaning of the particular word in Urdu precisely

| Urdu ID | English ID | English Word | Category | Concept | Example | Synsets |
|---------|------------|--------------|----------|--|---------|---------|
| 1 | | | N | کھانے کی شے جس کے ذریعے بدن کی پرورش ہوتی ہے | | کھانا |

8. بڑبڑ کر جانا، نگل جانا، قبضہ کر لینا، بضم کر لینا۔

"اب سننے میں آ رہا ہے کہ ناگیور کے ہونسلہ خاندان کی مرثہ ریاست کو بھی کھانا چاہتا ہے۔" (1946ء، جہانسی کی

رائی، 21)

9. ڈسنا، مٹانا، تکلیف دینا۔

"جو شخص اس لیے لوگوں سے بھیک مانگے اور اس ذریعہ سے اپنے مال کو بڑھائے تو یہ سوال قیامت کے دن اس کے

چہرے پر ایک زخم ہوگا اور دوزخ میں گرم پتھر بن کر اس کو کھائے گا۔" (1965ء، مشکوٰۃ شریف، 434:1)

10. جی یا جان کے ساتھ پریشان کرنا، تنگ کرنا۔

تس میں کیتا ہوں اپنے گھر جانے

حضرت شوق تم نہ جی کھانے (1798ء، دیوان سوز قدیم، 462)

11. (قسم یا حلف) اٹھانا، عید کرنا۔

تازہ غم کھایا کئے ہم وہ ہیں پاکیزہ مزاج

اور تم کھانے رہے چھوٹی قسم کھانی ہوئی (1884ء، آفتاب داغ، 142)

12. مطابقت رکھنا، ملنا۔

"یہ تمام عناصر مسلسل اور علی الترتیب مرتب نہیں ہون گئے بلکہ مجموعوں میں ہٹ جائیں گے اور ان میں بعض دوسروں

سے موافقت کھاتے معلوم ہوں گے۔" (1910ء، انبیا، نومبر، 208)

13. رشوت لینا، اثر قبول کرنا (سردی یا اوس وغیرہ کا)، چاٹنا، کھوکھلا کر دینا، اٹھنا، صرف ہونا، سننا

(گالی)۔ (نور اللغات)

Step 3: Exercise

| Urdu ID | English ID | English Word | Category | Concept | Example | Synsets |
|---------|------------|--------------|----------|--|---------|---------|
| 1 | 14972230 | Doomsday | N | مرنے کے بعد تمام انسانیت کے حساب کتاب کا دن | | قیامت |

Step 4

- Find out the English translation of the selected sense according to its determined POS tag in Urdu to English Dictionary e.g. the first sense of کھانا has following English translations available for Noun
- Eating, **food**, dinner, lunch

Step 5

- Look up the English translations of the selected Urdu word according to its determined POS tags in Princeton WordNet(PWN) version 2.1.

The noun dinner has 2 senses in PWN 2.1

1. (25) {07472733} <noun.food> dinner#1 -- (the main meal of the day served in the evening or at midday; "dinner will be at 8"; "on Sundays they had a large dinner when they returned from church")
2. (5) {08140714} <noun.group> dinner#2, dinner party#1 -- (a party of people assembled to have dinner together; "guests should never be late to a dinner party")

The noun lunch has 1 sense in PWN 2.1

1. (12) {07472083} <noun.food> lunch#1, luncheon#1, tiffin#1, dejeuner#1 -- (a midday meal)

The noun **food** has 3 senses in PWN 2.1

1. (34) {00020429} <noun.Tops> **food#1**, nutrient#1 -- (any substance that can be metabolized by an organism to give energy and build tissue)
2. {07453329} <noun.food> **food#2**, solid food#1 - - (any solid substance (as opposed to liquid) that is used as a source of nourishment; "food and drink")
3. {05739472} <noun.cognition> **food#3**, food for thought#1, intellectual nourishment#1 -- (anything that provides mental stimulus for thinking)

Step 6: English ID and English Word

- Pick up the English ID of the most relevant sense from Princeton WordNet

| Urdu ID | English ID | English Word | Category | Concept | Example | Synsets |
|---------|------------|--------------|----------|--|---------|---------|
| | 20429 | Food | N | کھانے کی شے جس کے ذریعے بدن کی پرورش ہوتی ہے | | کھانا |

Step 6: Exercise

| Urdu ID | English ID | English Word | Category | Concept | Example | Synsets |
|---------|------------|--------------|----------|---|---------|---------|
| | 14972230 | doomsday | N | مرنے کے بعد تمام انسانیت کے حساب کتاب کا دن | | قیامت |

Step 7: Example

- Write an example of the word in this particular sense with the same POS tag in simple and precise language to explain its concept

| Urdu ID | English ID | English Word | Category | Concept | Example | Synsets |
|---------|------------|--------------|----------|--|-------------------------------------|---------|
| | 20429 | Food | N | کھانے کی شے جس کے ذریعے بدن کی پرورش ہوتی ہے | کھانا کھانے سے بدن کی پرورش ہوتی ہے | کھانا |

Step 7: Exercise

| Urdu ID | English ID | English Word | Category | Concept | Example | Synsets |
|---------|------------|--------------|----------|---|--|---------|
| 1 | 14972230 | Doomsday | N | مرنے کے بعد تمام انسانیت کے حساب کتاب کا دن | قیامت کے دن اللہ تعالیٰ سب کا حساب کتاب کرے گا | قیامت |

Step 8: Synsets

- Find out the synonyms of the particular word in a synonym dictionary of Urdu e.g. قاموس مترادفات
- Confirm the concept of each synonym in Urdu Dictionary and write it

Step 8: Exercise

| Urdu ID | English ID | English Word | Category | Concept | Example | Synsets |
|---------|------------|--------------|----------|--|--|-------------------------|
| | 20429 | Food | N | کھانے کی شے جس کے ذریعے بدن کی پرورش ہوتی ہے | ابھی غذا ہے ابھی بدن کی پرورش ہوتی ہے | کھانا، خوراک، غذا |

| Urdu ID | English ID | English Word | Category | Concept | Example | Synsets |
|---------|------------|--------------|----------|---|--|---|
| 1 | 14972230 | Doomsday | N | مرنے کے بعد تمام انسانیت کے حساب کتاب کا دن | قیامت کے دن اللہ تعالیٰ سب کا حساب کتاب کرے گا | قیامت، روزِ جزاء، یوم الدین، یوم الحساب، یوم النشور، یوم نقان، یوم حشر، آخرت، محشر، عاقبت |

THANK YOU!
QUESTIONS?

References

- http://www.ling.uni-potsdam.de/~das/teaching/lexsem04/zugck_ho.pdf
- <http://wordnet.princeton.edu/>
- C. Fellbaum, "WordNet: An Electronic Lexical Database." MIT Press, Cambridge, Massachusetts, 1998.
- C. Fellbaum, M. Palmer, L. Delfs, S. Wolf, "Manual and Automatic Semantic Annotation with WordNet", 2001, Retrieved (11, 06, 2012). Available at: <https://nats-www.informatik.uni-hamburg.de/intern/proceedings/2001/naacl/mwnw/pdf/invitedPaper.pdf>
- F. Adeeba and S. Hussain, "Experiences in Building the Urdu WordNet", *IJCNLP*, 2011, Retrieved (11, 06, 2012). Available at: <http://www.cle.org.pk/Publication/papers/2011/UrduWordNet.pdf>
- T. Ahmed and A. Hautli, "Developing a Basic Lexical Resource for Urdu using Hindi WordNet", in proc. *CLT10*, Islamabad, 2010, Retrieved (11, 06, 2012). Available at: http://ling.uni-konstanz.de/pages/home/pargram_urdu/main/files/Ahmed_Hautli_CLT10.pdf