Structural Analysis of Linking Urdu WordNet to PWN 2.1

Ayesha Zafar, Afia Mahmood

Sana Shams & Dr. Sarmad Hussain

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

CLT14 | Center for Language Engineering (CLE), 2014

Outline

- Introduction
- Urdu WordNet
- Linkage of Urdu WordNet to Princeton WordNet 2.1
- Methodology
- Challenges and Proposed Solutions
- Un-Mapped Lexical and Cultural Senses
- Current Status
- Conclusions and Future Work

Introduction

- WordNet is a lexical database
  - Nouns, Verbs, Adj, Adv, are grouped into synsets representing a concept
  - Synsets are interlinked by means of conceptual, semantic and lexical relations

WordNet 2.1 Browser

File History Options Help

Search Word: fly

Searches for fly: Noun Verb Adjective

The noun fly has 5 senses (first 4 from tagged texts)

1. (6) {02168854} <noun.animal> fly#1 -- (two-winged insects characterized by active flight)
2. (1) {04360018} <noun.artifact> tent-fly#1, rainfly#1, fly sheet#1, fly1#2, tent flap#1 -- (flap consisting of a piece of canvas that can be drawn back to provide entrance to a tent)
3. (1) {03337531} <noun.artifact> fly#3, fly front#1 -- (an opening in a garment that is closed by a zipper or by buttons concealed under a fold of cloth)
4. (1) {00126081} <noun.act> fly#4, fly ball#1 -- ((baseball) a hit that flies up in the air)
5. {03337708} <noun.artifact> fly2#5 -- (fisherman's lure consisting of a fishhook decorated to look like an insect)

The verb fly has 14 senses (first 9 from tagged texts)

1. (33) {01922384} <verb.motion> fly#1, wing#1 -- (travel through the air; be airborne; "Man cannot fly")
2. (9) {01824853} <verb.motion> fly2#2 -- (move quickly or suddenly; "He flew about the place")
3. (5) {01923074} <verb.motion> fly1#3, aviate#1, pilot#1 -- (fly a plane)
4. (3) {01439135} <verb.contact> fly#4 -- (transport by aeroplane; "We fly flowers from the Caribbean to North America")
5. (2) {01830701} <verb.motion> fly7#5 -- (cause to fly or float; "fly a kite")
6. (2) {01822189} <verb.motion> fly8#6 -- (be dispersed or disseminated; "Rumors and accusations are flying")
7. (2) {00539899} <verb.change> fly1#7 -- (change quickly from one emotional state to another; "fly into a rage")
8. (1) {02054741} <verb.motion> fly6#8, fell9#2, vanish5#3 -- (pass away rapidly; "Time flies like an arrow"; "Time fleeing beneath him")
9. (1) {01823023} <verb.motion> fly5#9 -- (travel in an airplane; "she is flying to Cincinnati tonight"; "Are we driving or flying?")
10. {02121116} <verb.perception> fly#10 -- (display in the air or cause to float; "fly a kite"; "All nations fly their flags in front of the U.N.")
11. {02057156} <verb.motion> flee#1, fly3#11, take flight#1 -- (run away quickly; "He threw down his gun and fled")
12. {01830488} <verb.motion> fly4#12 -- (travel over (an area of land or sea) in an aircraft; "Lindbergh was the first to fly the Atlantic")
13. {01390557} <verb.contact> fly1#13 -- (hit a fly)
14. {00150498} <verb.change> vanish1#5, fly#14, vaporize4#4 -- (decrease rapidly and disappear; "the money vanished in las Vegas"; "all my stock assets have vaporized")
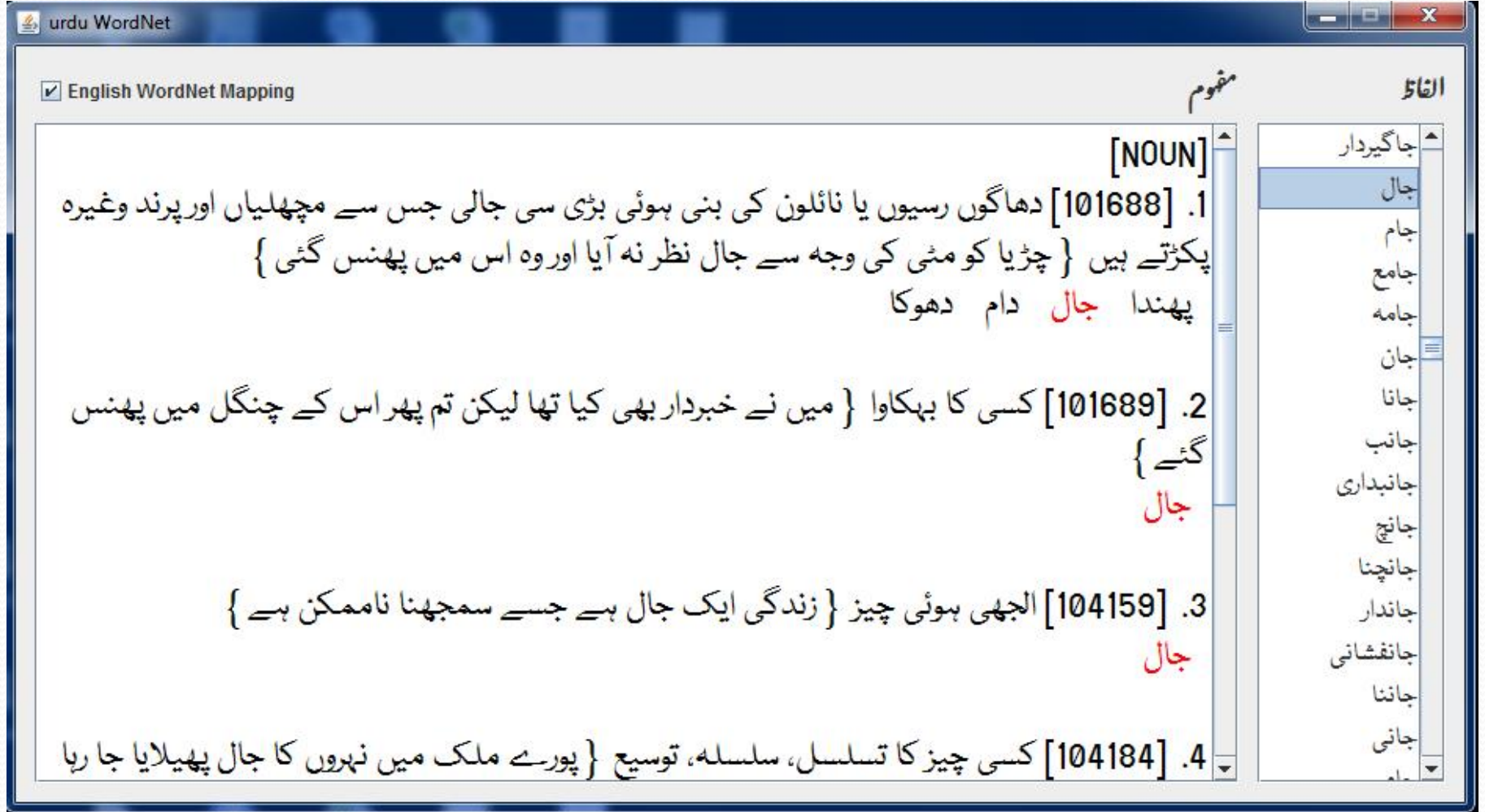
The adj fly has 1 sense (no senses from tagged texts)

1. {00099145} <adj.all> fly#1 -- ((British informal) not to be deceived or hoodwinked)

Urdu WordNet

- Urdu WordNet (http://cle.org.pk/clestore/urduwordnet.htm) is the first semantic dictionary of Urdu
  - It's a work in progress, and currently it has 5058 Urdu senses/ concepts
  - Each Urdu sense has
    - POS definition
    - Unique synset ID
    - Definition/ Concept
    - Synset
    - Example

Example: Layout of Urdu WordNet

Linkage of Urdu WordNet to PWN 2.1 - Process

- Enter and verify the Urdu Word and its senses

- Look-up the meaning in dictionaries

- Generate at least 3 candidate alignment terms

- Select the term by carefully analyzing its meaning as well as complete POS category

- Map the Urdu sense with the selected PWN sense ID

Linkage of Urdu WordNet to PWN 2.1 - Process

Urdu Word	امن	امن	امن
Urdu POS	N	N	N
Urdu Concept	چین اور اطمینان سے بھرپور ماحول یا جگہ (پر کے ساتھ)	جنگ کی ضد	سکون کی حالت
Urdu Sentence	بہت جلد اسے دشمنوں سے نجات کی صورت میں ایک پر امن جگہ مل گئی	اس بار جنگ وجدل سے کام لینے بجائے ہم امن کا پیغام لے کر جائیں گے	امیر شہر کے آتے ہی لوگ امن سے رہنے لگے
Candidate Terms	peace, repose, reconciliation	peace, harmony, accord	Tranquility, calm, serenity
Selected Eng Word	peace	peace	tranquility
English Concept	the absence of mental stress or anxiety	the state prevailing during the absence of war	an untroubled state; free from disturbance
POS in PWN	noun.feeling	noun.state	noun.state
Eng Sense ID	07413685	13784195	13783084

Alignment Challenges
&
Proposed Solutions

1. Morphological Issue:
Causative difference of Urdu and English

- Urdu has morphological devices such as inflection, that change verbs into their causative forms, with infixes like لا (lā-)and وا (vā-)

متعدی المتعدی (di-transitive verb)	متعدی (transitive verb)	لازم (root verb)
سلوانا	سلانا	سونا
بجوانا	بجانا	بجنا

1. Morphological Issue:
Causative difference of Urdu and English

- سونا maps on sleep ({13838748} <noun.state>)

- سلانا (to make someone sleep) could not be found in PWN
- سلوانا (to make someone put someone to sleep) could not be found in PWN

2. Syntactic Issue: Complex Predicates

- Two types of complex predicates couldn't be mapped
  - افشا کرنا (to disclose).......................... N+V
  - اثرانداز ہونا (to influence)....................Adj+V

- They use light verb to give complete meanings and syntactically and semantically they behave like single constituents

3. Semantic Issue: i) Single Urdu concept for multiple PWN concepts

بچپن
کم سن ہونے کی کیفیت، نو عمری

Noun.time	noun.state
{14948030}	{14235403}
(the time of person's life when they are a child)	(the state of a child between infancy and adolescence)

3. Semantic Issue: ii) Multiple Urdu concept for single PWN concepts

بدکنا (scared) جانور کا ڈر کر یا بگڑ کر بھڑکنا (animals scared and retreats)	بدکنا (scared) آدمی کا یکایک کسی سے ڈر کر بدگمان ہونا، الگ ہو جانا (suddenly a man gets scared to be skeptical)

Scared
{01762161} <verb.emotion> cause fear in

3. Semantic Issue: iii) Difference in Personal Relationship

- **Blood relations**
  - Urdu language carries different terms for blood relations
  - nephew {10198902} <noun.person> in PWN is used as a son of your brother or sister
  - whereas in UWN بھانجا means sister's son and بھتیجا is used for brother's son

3. Semantic Issue: iii) Difference in Personal Relationship

- **Relations with In-laws**
- Urdu lexicalizes the distinction between the blood relations of husband and wife:
  - سالا is used for wife's brother
  - جیٹھ elder brother of husband and دیور younger brother of husband
- In PWN only one sense exist: Brother-in-law {09731744} <noun.person> a brother by marriage

3. Semantic Issue: iii) Difference in Personal Relationship

- **Maternal and Paternal relations**
  - English does not have specific concepts for relationships
  - چچا (younger paternal uncle), تایا (elder paternal uncle), ماموں (maternal uncle)

- Nearest corresponding English sense in PWN:
  - uncle {10575646} <noun.person> -- (the brother of your father or mother; the husband of your aunt)

3. Semantic Issue: iv) Differences in names of utensils, fruits, etc.

- Certain kitchen utensils depicts indigenous culture

- ڈونگا
- لکڑی کا بڑا چمچہ، کونڈا، برتن جس میں شوربا وغیرہ چھنتے ہیں، پیالے کی شکل کا چھوٹا ظرف

- PWN gives a general concept of utensil i.e. an implement for practical use (especially in a household)

3. Semantic Issue: iv) Differences in names of utensils, fruits, etc.

- There are many fruit names which are culture specific

- کیری (unripe mango fruit) is commonly used in Urdu

- For example, کیری can be linked to English synset mango

Proposed Solution: Construction of hypernymy linkage

Un-mapped Lexical and Cultural Senses

- **Cultural specific vegetables**
  - Only exist in Urdu that's why cannot be mapped
  - e.g. بتھوا ، ساگ

- **Semantic orientation of borrowed words**
  - پوسٹ (post/ any office or rank), is a borrowed word from English
  - It gives different sense in Urdu

- **Literal Concepts**
  - Many words in Urdu language are based on stereotypes and culturally-inherited associations, they remain unmapped as no parallel senses exist

Words	Concept	Example
بھوننا	بندوق سے گولیاں مار کر قتل کر دینا	فوج نے ایک ہی حملہ کیا اور دشمن کے کئی سپاہی گولیوں سے بھون دیے
پھڑکنا	غیر معمولی حرکت یا جنبش	صبح سے ہی اس کی آنکھ پھڑک رہی تھی

Current Status of Work

- **3526** Urdu senses reviewed
- **1829** aligned to PWN 2.1.
- **1,403** remained unmapped

Total count of Nouns, Adjective and Verbs from UWN	
Total number of Nouns	1002
Total number of Adjectives	872
Total number of Verbs	249

Conclusions and Future Work

- 1829 senses are aligned to PWN 2.0, further research needs to be conducted in
  - Hypernymy relationship development
  - Urdu VerbNet development
- Automatic and Semi-Automated approaches for alignment

Acknowledgement

This work has been supported through a grant from

German Research Exchange Program (DAAD)

in collaboration with

University of Konstanz, Germany

Thank you

Sana.Shams@kics.edu.pk

References

- P. Vossen, "EuroWordNet: a Multilingual Database for Information Retrieval", workshop on Cross-language Information, Zurich. Available at: http://www.ercim.eu/publication/ws-proceedings/DELOS3/Vossen.pdf

- E. Pianta, L. Bentivogli, C. Girardi, "MultiWordNet: Developing an Aligned Multilingual Database", In Proceedings of the *First International Conference on Global WordNet*, Mysore, India, 2002. Available at: http://multiwordnet.fbk.eu/paper/MWN-India-published.pdf

- J. Ramanand, A. Ukey, B. Singh, P. Bhattacharyya, "Mapping and Structural Analysis of Multi-lingual Wordnets", In Proceedings of *IEEE,* Bombay, 2007. Available at: http://www.cse.iitb.ac.in/~pb/papers/IEEE_article_IIT_Bombay_vFeb08.pdf

- C. Fellbaum, M. Palmer, L. Delfs, S. Wolf, "Manual and Automatic Semantic Annotation with WordNet", In Proceedings of *NAACL,* Pittsburgh, 2001. Available at: https://nats-www.informatik.uni-hamburg.de/intern/proceedings/2001/naacl/mwnw/pdf/invitedPaper.pdf

- J. Daude, L. Padro, G. Rigau, "Mapping Wordnets Using Structural Information." 38th Annual Meeting of the Association for Computational Linguistics, 2000. Available at: http://aclweb.org/anthology//P/P00/P00-1064.pdf

- J. Saraswat, S. Ripple, P. Goyal, P. Bhattacharyya, "Hindi to English Wordnet Linkage: Challenges and Solutions". Available at: http://www.cfilt.iitb.ac.in/wordnet/webhwn/IndoWordnetPapers/04_iwn_Hindi%20to%20English%20Wordnet%20Linkage%20-%20Challenges%20and%20Solutions.pdf

- A. Zafar, A. Mahmood, F. Abdullah, S. Zahid, S. Hussain, and A. Mustafa, "Developing Urdu WordNet Using the Merge Approach ", in the *Proceedings of Conference on Language and Technology* 2012 (CLT12), Lahore, Pakistan. Available at: http://www.cle.org.pk/clt12