Development of Treebank from parallel corpora using Factory model of Anusaaraka

8-3-16/11.30 a.m.

1. See whether P layer (Manual translation) is aligned with A layer (English Sentence). If it is correct then phrasal alignment tool is working fine.
2. Check N th layer (parser alignment) alignment with A th layer. If it is not aligned properly there can be three reasons for that
   1. Phrasal alignment algorithm written by Manju mam is going wrong
   2. Stanford parser for english is giving wrong output.
   3. Hindi rule based parser (Roja ma’am) giving wrong output.

3. Debug all these a,b,c cases.

Scope of my thesis

Case a:

This implementation is done by Manju mam in clips and i need to see where it is going wrong (if errors). I can take this task of phrasal alignment and create an algorithm and implement it in whatever lang I want.

Case b:

Here I can find out what the patterns of error making by stanford parser. And I can write/modify post-processor for stanford parser which will give me correct output.

Study of stanford error parser- Berkeley paper my VC sir

Case c:

I need to report the errors to roja mam, she will check the rules written for formulation of hindi parser.

\*\*Development of a functional/? language for common users where they can write their own program to solve their daily work. (Prof. Noori)

\*\* Goal : Development of hindi parser using any functional programming language. (Just like => Amba mam designed a sanskrit parser in ocaml ) : Study upto functors, monads

Study haskell and take time from mam and VC sir to discuss on it.

Difficulties I am facing while debugging:

1. **GUI**: All html/dat files I am checking are not arranged in user-friendly manner. I need to remember which file is for what purpose. And where is the location of that file. Once I knew what files i need to check I should have well arranged GUI where I can just click on certain things and see o/p.
   1. List out which different files I need to check if i want to check parser o/p of hindi and english sentence parallely
   2. Design a GUI which will arrange this very properly.
   3. **Change relation.dat to tree format** : Checking output of stanford/any parser (dependency) in user friendly manner.
2. Rule based parser is giving multiple relations for the same word pairs, so phrasal alignment doesn’t work in this case. => Temporary Solution : Use stable version of rule based parser , take it from Roja mam and in phrasal alignment use it

Permanent Solution => improve hindi parser (roja mam), understanding and maintenance of hindi rule based parser. How? - Give 5-10 multiple outputs of parser relation to soma mam and she will correct it by writing a rules in clips, add it to hindi rule based parser (roja mam). Understanding git.

=================================

Temp directory having output of all anusaaraka : $HOME\_anu\_tmp

Stanford relations : relations.dat

Hindi relations : hindi\_parser\_relation\_wrd.dat

Hindi words, ids : manual\_word.dat

anu\_clips/sd\_relation\_rules.clp => stanford relation mapping to paninian relations

misc/SMT/dep-parser/lex => mapping of karak relations to paninian relations, eg k1-kriya-sambhandhi

1-5-17

Task : Deriving hindi parse from stanford dependency english parser and phrasal alignment.

Requirement

1. The parallel sentences should be very close to each other
2. English parse should be correct
3. Alignment must be correct
4. <http://corenlp.run/> check eng parse here
5. <http://ltrc.iiit.ac.in/full_analyzer/hindi/> hindi UD tree
6. Paninian tree of hindi can be seen from hindi rule based parser which is not upto the mark yet.
7. The stanford relations and Hindi UD tree similarity.
8. The stanford relations and Paninian relations similarity.

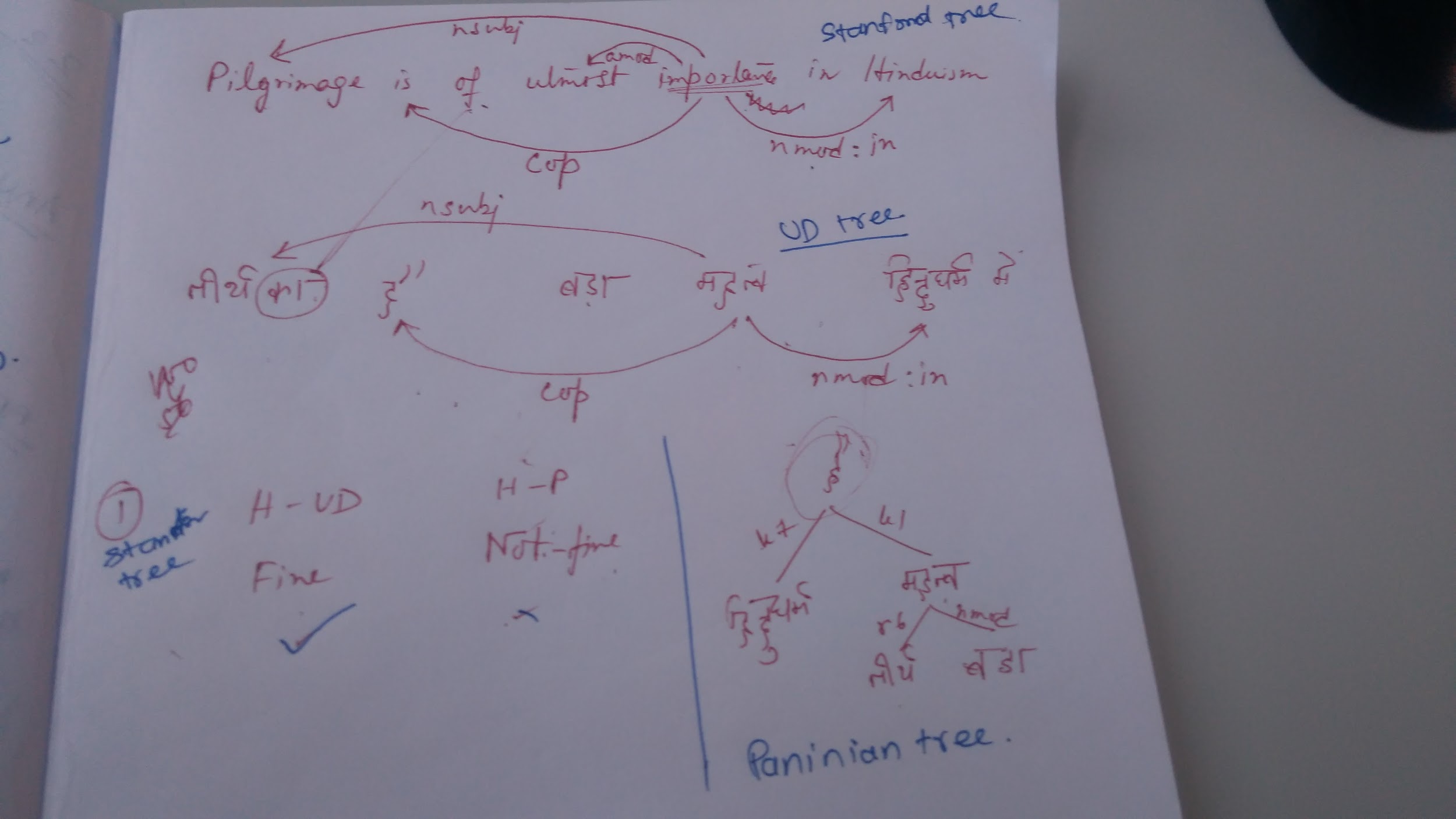
ask to manju mam- please send me readme from which i can know that which file contains what in phrasal alignment or atleast create two different directories for saving eng info n hindi info files

2.2

हिंदू धर्म में तीर्थ का बडा महत्व है .

Pilgrimage is of utmost importance in Hinduism.

1. ??? Word level left over Words:: 1 हिंदू 2 9 . 10 : but still this word is aligned in phrasal how? And in parsing hindu is missing



<https://docs.google.com/spreadsheets/d/1L4upe420dyuLW_p_4Fr55usM_-dbDp2iYoTn_L3ksNU/edit?usp=sharing>

8-5-2017

The work which I done in above file was wrong…

1. Every word must be aligned in phrasal alignment tool
2. Check parse of english sentence in stanford corenlp.
3. Project all the relations from stanford parser to hindi chunks as they are aligned
4. By your own knowledge check whether that relations are correct according to hndi-UD parsing
5. If yes fine else take out the exceptions