

Lecture 6: HW2 Review, Corpus Linguistics

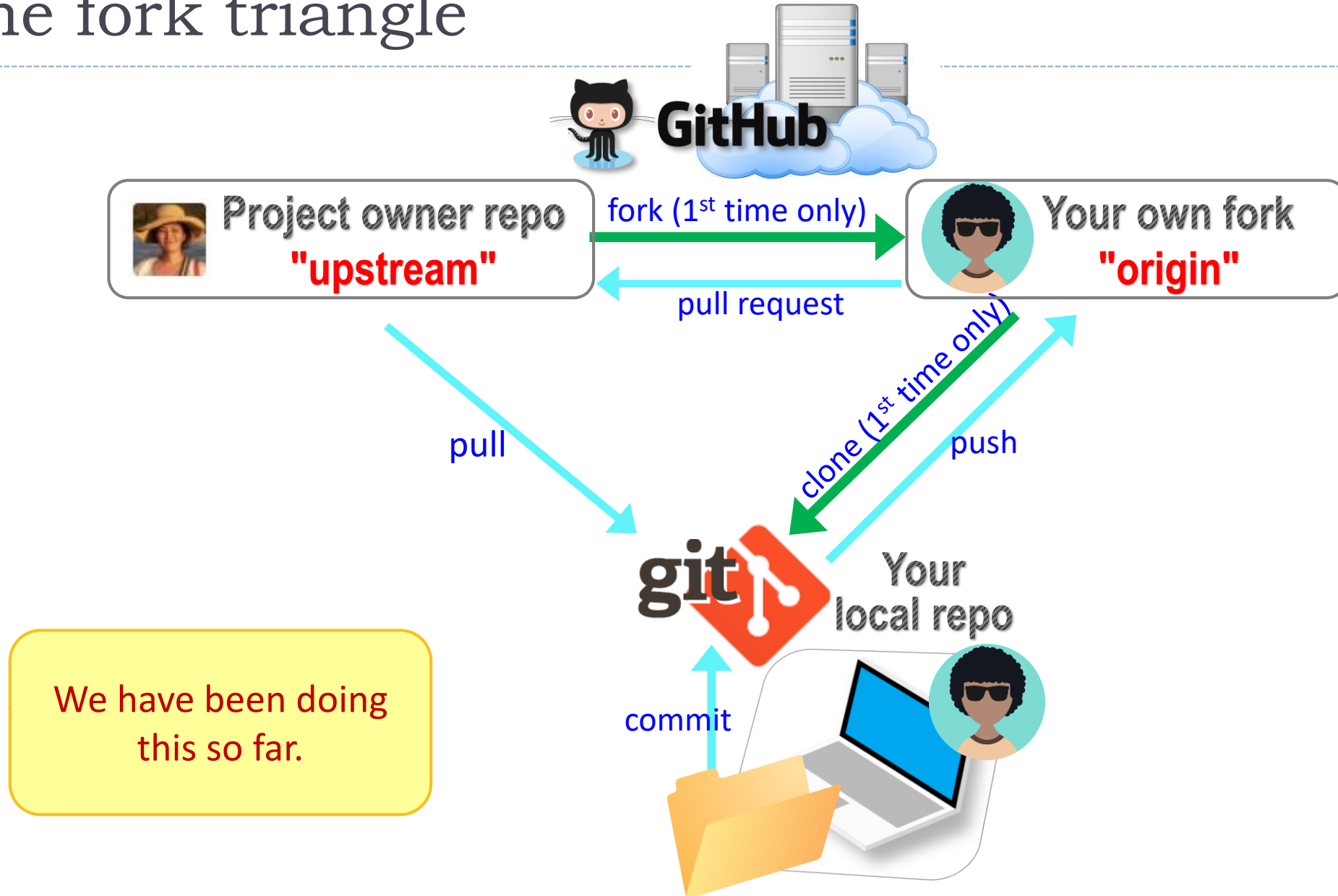
LING 1340/2340: Data Science for Linguists

Na-Rae Han

Corpus linguistics

- ▶ Review Gries & Newman (2013) "Creating and using corpora"
 - ◆ <https://naraehan.github.io/Data-Science-for-Linguists-2019/resources#corpus>
- ▶ What corpus resources are cited?
- ▶ What tools?
- ▶ Which important concepts?

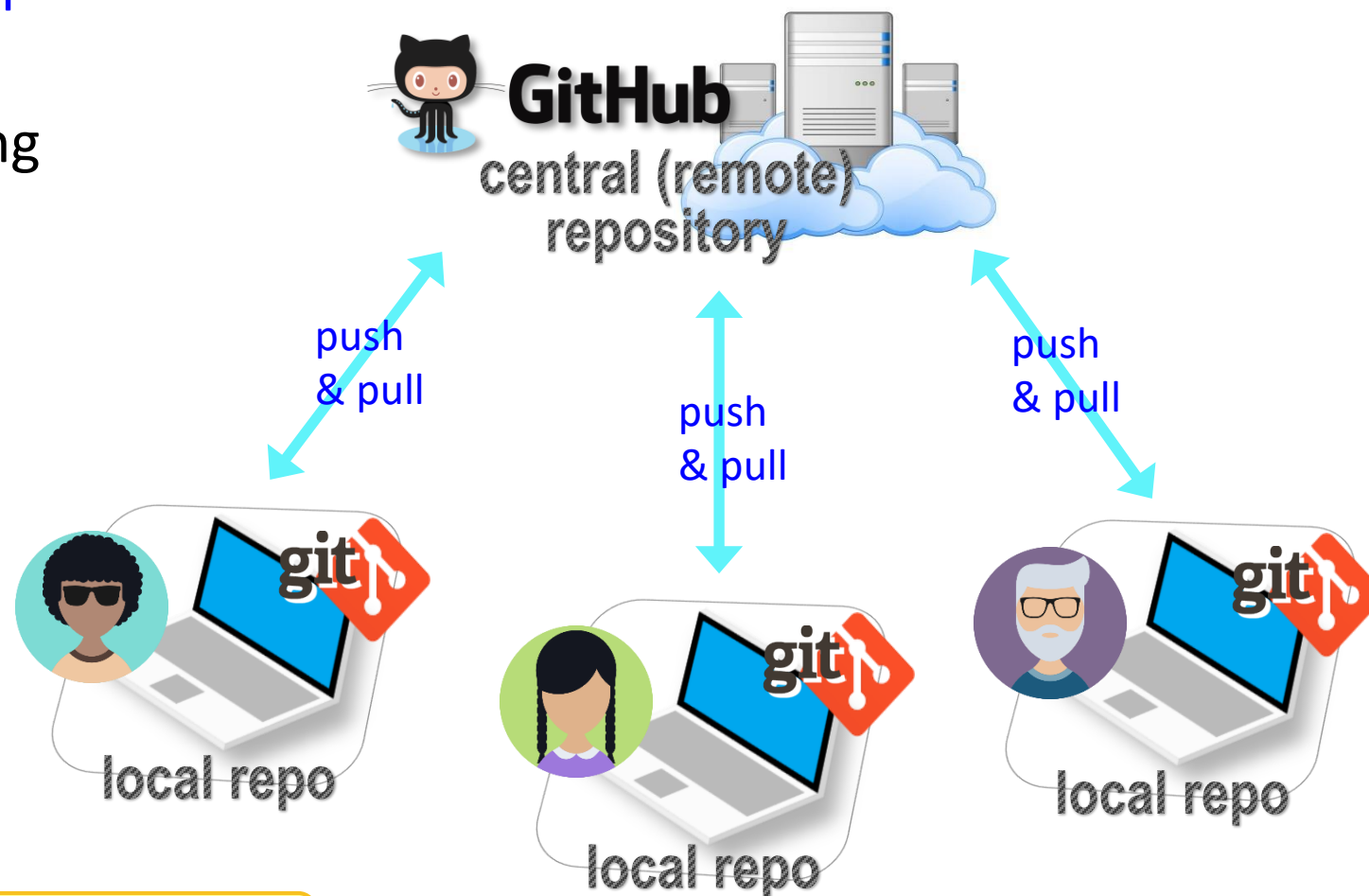
The fork triangle



GitHub: a *social*, remote repository

▶ GitHub also works as a **central remote repository** among a group of **collaborators** working on a shared project.

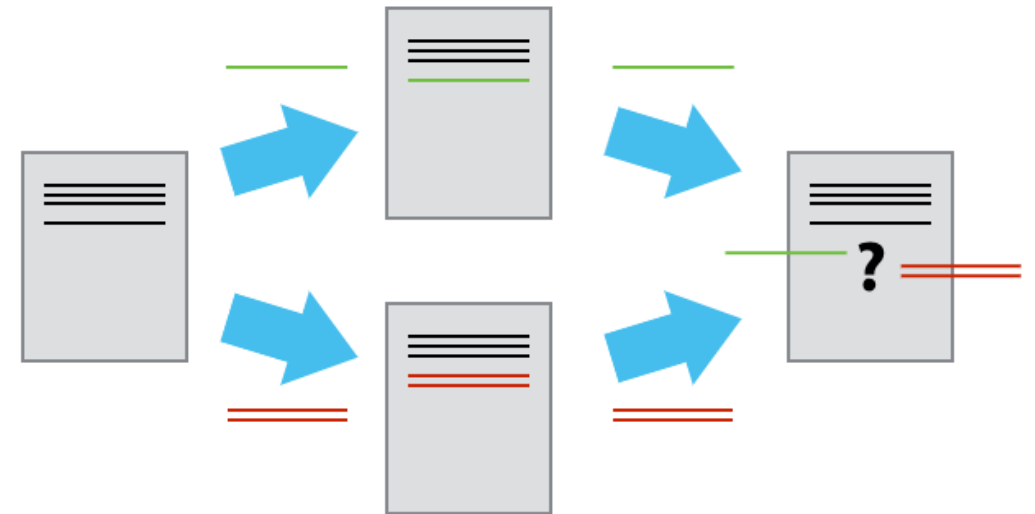
- ◆ Everyone works on their own *local* copy of the repository, making changes.
- ◆ Git is able to keep track and merge changes submitted by everyone.
- ◆ Everyone is an **equal collaborator** with push (=write) access.



We are now ready!

Introducing... "Class-Plaza"

- ▶ Public.
- ▶ Everyone is listed as a "collaborator".
 - ◆ Meaning, everyone has push access.
 - ◆ No need to fork: pull and push directly.
 - ◆ We will also truly collaborate: **edit shared files**.
- ▶ This means: **CONFLICTS**
 - ◆ Na-Rae's tutorial on Git conflicts:
 - ◆ https://github.com/mcdonn/LSA2019-Reproducible-Research/blob/master/linking_git_and_github.md#conflicts



When there is a conflict



- ▶ After you pull, Git changes your file, which then looks like:

```
<<<<<<< HEAD
There was copyright information here.

=====
>>>>>>> c954b23f86b629c569223e2c0c38e32a0d870d22
"RTlexdec", "RTnaming", "Familiarity", "Word", "AgeSubject", "WordC
tenFrequency", "WrittenSpokenFrequencyRatio", "FamilySize", "Deri
y", "InflectionalEntropy", "NumberSimplexSynsets", "NumberComplex
thInLetters", "Ncount", "MeanBigramFrequency", "FrequencyInitialD
elV", "ConspelN", "ConphonV", "ConphonN", "ConfriendsV", "Confriend
ConffN", "ConfbV", "ConfbN", "NounFrequency", "VerbFrequency", "CV"
Frication", "Voice", "FrequencyInitialDiphoneword", "FrequencyIni
lable", "CorrectLexdec"
```

You must **manually edit this file** and tidy it up.
(== resolve conflict)

A GitHub race: our favorite animals

Activity
10 minutes



1. Everyone was already added to the repo as a collaborator.
2. Clone the repo to your laptop.
3. Edit "animals.md", add your line.
4. Do your usual local git routine: adding, committing.
5. Try pushing. It is likely you have a conflict (someone else pushed in the meantime) and Git tells you to pull first.
6. Pull to receive the new updates.
7. Open "animals.md". Resolve conflict.
8. Go back to step 4. Hope you were quick enough this time!

Wrapping up

- ▶ To-do #6 out: corpus resources
 - ◆ Make sure to properly handle conflicts!

- ▶ Your project
 - ◆ Feedback will be forthcoming. You should make a good progress in the next couple of weeks.
 - ◆ Copyright & licensing issues -- start thinking about them.

- ▶ Na-Rae's Monday office hours
 - ◆ Changed to 4:30 -- 5:30pm