I can answer for the Twitter data. You can gather plenty of French language tweets by one of two methods:

- Live Stream. In this case, you can pass this URL: https://stream.twitter.com/1.1/statuses/sample.json?language=fr and you will receive tweets as the standard JSON dictionary. An example python code to download a portion of the public stream is here. In this case, you let the code run and stop it once you have collected a decent sized corpus.

- Search for French language tweets. To do this programatically, you can use a code like this but replace the search terms with lang:fr. In this case, you search back for a period up to one week for statuses (aka tweets) that are marked as French language.

Notes:

- For both of these methods, you have to authenticate as a developer first (link).

- The language is based on Twitter's algorithm for recognizing language - and is not perfect.

- It's against the Twitter Terms of Service to share raw twitter data, but you can easily collect GBs of data by one the above methods.

From:
https://wiki.centrale-med.fr/informatique/ - **WiKi informatique**

Permanent link:
https://wiki.centrale-med.fr/informatique/public:twitter_data

Last update: **2017/03/26 10:47**