

Диалоговые датасеты на русском:

1. Субтитры
<http://subs.com.ru> (26330)
(12846 фильмов)
ссылки на другие ресурсы (<http://subtitry.ru/links/>)
2. Драматургия
<http://www.theatre-library.ru> - всего более 10500 произведений от более чем 3400 авторов.
Существенным недостатком является отсутствие единообразия в описании действия и связанные с этим сложности при обработке.
Русдракор (<https://github.com/lehkost/RusDraCor>) 69 драм в формате TEI-P5
Современная драматургия (75 произведений) <http://lib.ru/PXESY/>
3. Twitter
API позволяет скачивать твиты (так пишет этот исследователь <http://study.mokoron.com/>).
Она использовала twitter для сбора датасета положительных и отрицательных отзывов).
4. Сценарии
<http://screenwriter.ru/>
<http://www.lavkascenariiev.com>
<http://www.ezhe.ru/data/vgik/library.html>
<http://www.kinomania.ru/scripts>