



# J-LIWC2015の紹介

**五十嵐 祐**（名古屋大学）

日本心理学会第85回大会（オンライン）チュートリアル・ワークショップ

J-LIWC2015による日本語テキスト解析

2021年9月2日

# 出典

- Igarashi, T., Okuda, S., & Sasahara, K. (2021, August 28). Development of the Japanese Version of the Linguistic Inquiry and Word Count Dictionary 2015 (J-LIWC2015).

<https://doi.org/10.31234/osf.io/5hq7d>

J-LIWC2015

1

**Development of the Japanese Version of  
the Linguistic Inquiry and Word Count Dictionary 2015 (J-LIWC2015)**

Tasuku Igarashi<sup>1</sup>, Shimpei Okuda<sup>2</sup>, and Kazutoshi Sasahara<sup>3</sup>

<sup>1</sup> Graduate School of Education and Human Development, Nagoya University

<sup>2</sup> Graduate School of Informatics, Nagoya University

<sup>3</sup> School of Environment and Society, Tokyo Institute of Technology

**Author Note**

Tasuku Igarashi  <https://orcid.org/0000-0001-9432-4425>

Kazutoshi Sasahara  <https://orcid.org/0000-0003-2572-0648>

Pennebaker Conglomerates, Inc. owns all rights to the J-LIWC2015 dictionary. All royalties given to Tasuku Igarashi and Kazutoshi Sasahara for the commercial use of the J-LIWC2015 dictionary will be donated to Nagoya University and Tokyo Institute of Technology.

# 自然言語処理による心的状態の解析

- 「辞書」ベースの分析

- 分析対象のテキストにおける、特定の評価軸（カテゴリ）に紐付いた「辞書語」の出現頻度をカウント（カウントベースの分析とも呼ばれる）
- 言語行動に反映される書き手や話し手の心的状態を定量的に分析
- 問題点1：辞書の開発には多大なコストがかかる
- 問題点2：多くの辞書は、感情を中心とする少数のカテゴリのみを扱っている

# 一般的な辞書ベースの分析 (イメージ)

(①前処理でテキストを単語に分割)

彼は親切だ

ポジティブ感情

(②辞書語とのマッチング、カウント)

カテゴリ	割合
ポジティブ感情	1/4

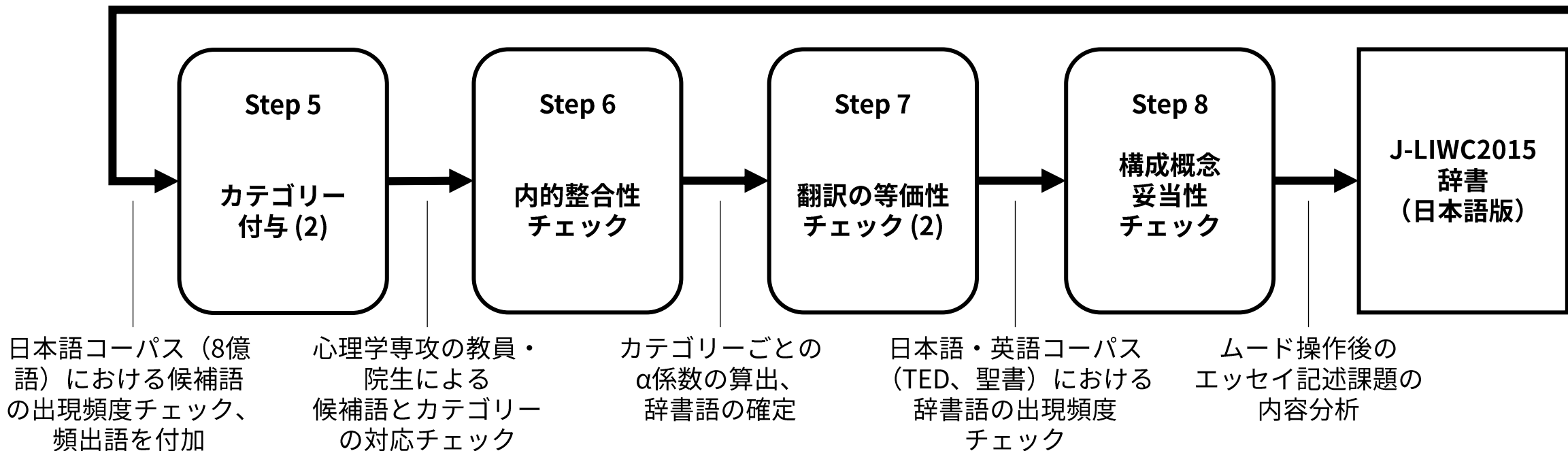
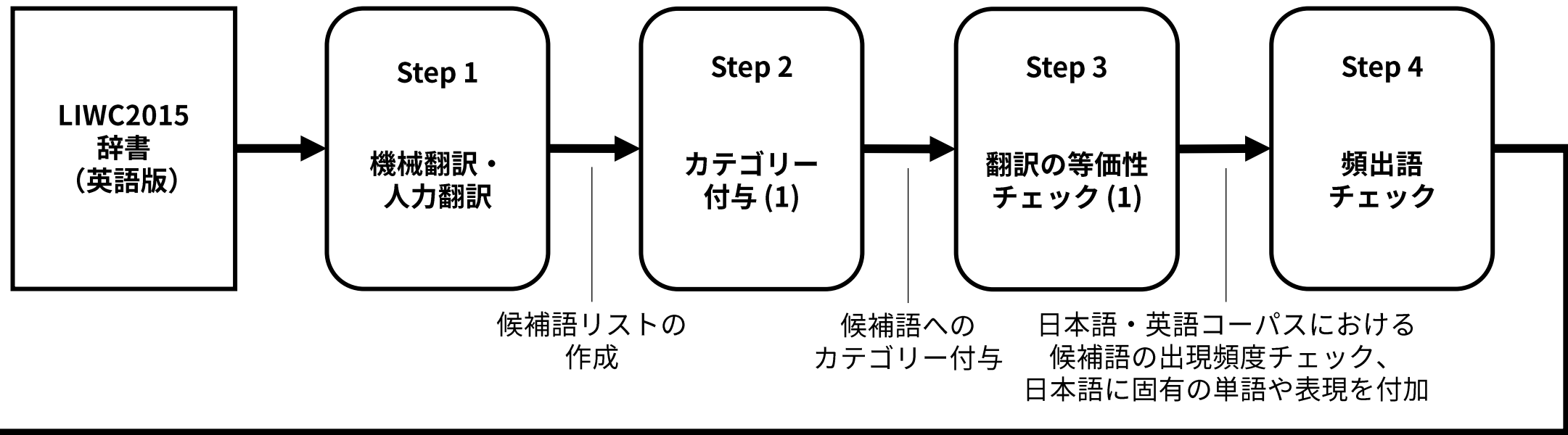
# LIWC2015 (Pennebaker, Boyd, Jordan, & Blackburn, 2015)

- **多様な言語カテゴリーと心理的カテゴリーを含む辞書**
  - 正式名称：Linguistic Inquiry and Word Count Dictionary 2015
    - 英語では「Luke」と発音 (Chung & Pennebaker, 2012)
  - 初版は1996年、改訂版は2001年、2007年、2015年にリリース
  - LIWC2015は、英語版の辞書と分析ソフトウェアが一体となっている
  - 多様な心的プロセスを反映する抽象的なカテゴリーを含む：すべてのカテゴリーについて、心理学者による信頼性・妥当性検証を行っている
- **辞書ベースの分析の実質的な世界標準**
  - 英語版がオリジナル：これまでに、ドイツ語、中国語（簡体字・繁体字）、スペイン語、ロシア語、アラビア語、フランス語、イタリア語、ポルトガル語、セルビア語、ルーマニア語、トルコ語に翻訳済（旧版含む）
  - Pennebaker et al. (2015) の論文の引用回数：3191件（Google Scholar, 2021/9/1現在）

# J-LIWC2015 (Igarashi, Okuda, & Sasahara, 2021)

## • LIWC2015の日本語版辞書

- 総単語数11,600、カテゴリー数69
- 2021年8月リリース、開発期間5年（2016年～2021年）
- Pennebaker Conglomerates, Inc. とライセンス契約
- 英語版、ドイツ語版の開発手順を踏襲
- MeCab（形態素解析エンジン）によるテキストの前処理後、LIWC2015ソフトウェアでの分析、プログラミング言語（Python, R etc.）での分析が可能
- MeCabの出力を利用して、基本的な品詞情報を後処理で補完可能
- 非商用利用は無料、商用利用はReceptivitiに要問合せ
- **辞書ファイルのダウンロード・利用には、LIWC2015ソフトウェアのライセンスが必要（PythonやRで分析する場合も）**



J-LIWC2015の開発フロー  
(Igarashi, Okuda, & Sasahara, 2021, Figure 1を改変)

言語次元	心理的プロセス	生物学的プロセス	インフォーマル
機能語	感情	身体	罵倒
代名詞	ポジティブ感情	健康	ネットスラング
人称代名詞	ネガティブ感情	性	うなずき
一人称単数	不安	摂取	間投詞
一人称複数	怒り	動因	フィラー
二人称	悲しみ	つながり	
三人称単数	社会的（相互作用）プロセス	達成	
三人称複数	家族	社会的地位・権力	
不定代名詞	友人	報酬	
格助詞	女性	リスク	
助動詞	男性	相対性	
副詞	認知プロセス	動作	
接続詞	洞察	空間	
否定詞	原因	時間	
その他の文法	不一致	個人的な事柄	
動詞	あいまいさ	仕事・学業	
疑問詞	確かさ	趣味・余暇	
数詞	差別化	家	
数量詞・助数詞	知覚プロセス	金銭	
形容動詞	視覚・知覚	宗教	
連体詞	聴覚	死	
	感覚（触覚・味覚・嗅覚）		
	摂取		



# なぜ心理学者によるチェックが必要か

- LIWC2015のカテゴリーの多くは、理論的構成概念を反映
  - 例えば、「報酬」カテゴリーには、「近寄る」、「促進」、「前向き」といった語が含まれる
    - これは、Gray (1970) の強化感受性理論（報酬への接近感受性）と関連すると考えられる：心理学の知識がなければ、なぜこれらの語が報酬と関連するのか判断できない
- 日本語版の開発にあたっては、社会心理学、認知科学、発達心理学、臨床心理学を専攻する大学院生のサポートを得て、カテゴリーの内容的妥当性を確認

# J-LIWC2015での分析（イメージ）

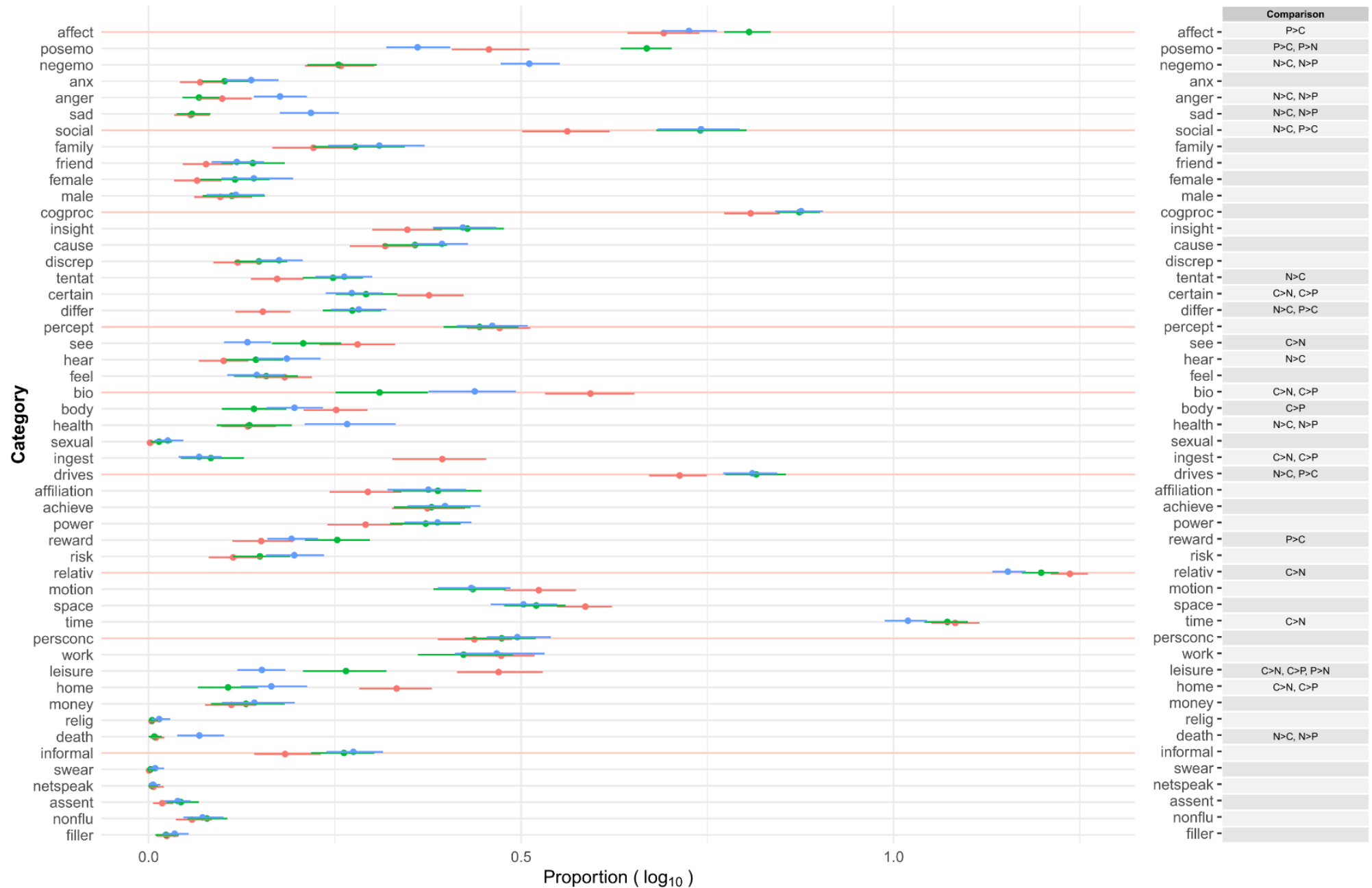
（①前処理でテキストを単語に分割）

彼は親切だ

- 機能語
- 代名詞
- 人称代名詞
- 三人称単数
- 社会的（相互作用）プロセス
- 男性
- 機能語
- 格助詞
- 感情
- ポジティブ感情
- 社会的（相互作用）プロセス
- 友人
- 動因
- つながり
- 機能語
- 助動詞

（②辞書語とのマッチング、カウント）

カテゴリ	割合
機能語	3/4
代名詞	1/4
人称代名詞	1/4
三人称単数	1/4
格助詞	1/4
助動詞	1/4
感情	1/4
ポジティブ感情	1/4
社会的（相互作用）プロセス	2/4
友人	1/4
男性	1/4
動因	1/4
つながり	1/4



ムード操作後のエッセイ内容の分析  
(Igarashi, Okuda, & Sasahara, 2021, Figure 3)

# 心理学領域での応用可能性

- 臨床場面における面接内容の定量化
  - SNSの投稿内容の国際比較
  - ペアにおける発話傾向の一致度の検討
  - 発話内容と心理的健康との関連
  - 非意識的プロセスと言語行動との対応
- …など

# お気づきの点は

- GitHubでIssueを作成してください（日本語可）
  - サンプルスクリプトのバグ
  - 自身で作成した前処理&後処理&分析スクリプト（Colab）の共有依頼
  - 他言語（Rなど）のスクリプト（Colab）の共有依頼

The screenshot shows the GitHub repository page for `tasukuigarashi / j-liwc2015`. The repository has 2 watchers, 0 stars, and 0 forks. The main navigation bar includes links for Code, Issues, Pull requests, Actions, Projects, Wiki, Security, and Insights. The repository is currently on the `master` branch. A recent commit by `tasukuigarashi` is shown, titled "Merge branch 'master' of github.com:tasukuigara...", with a timestamp of 24 minutes ago. The commit history table lists the following files and their commit times:

File	Commit Message	Time
jpa2021ws	sample file upload for JPA2021	25 minutes ago
LICENSE	first commit	3 years ago
README.md	Merge branch 'master' of github.com:tasukuigara...	24 minutes ago
user_dict.dic	user dictionary	12 days ago

The `README.md` file is selected, showing the title "J-LIWC2015" and an "Overview" section. The overview text states: "This repository explains how to analyze Japanese texts by J-LIWC2015. J-LIWC2015 is a Japanese-translated version of the LIWC2015 dictionary (Pennebaker, Booth, Jordan, & Blackburn 2015). Psychometric properties of the J-LIWC2015 dictionary are introduced".

On the right side of the repository page, there are sections for "About" (Japanese version of LIWC2015), "Readme", "MIT License", "Releases" (No releases published, Create a new release), and "Packages" (No packages published, Publish your first package).