

LIWC

Daichi Sakushima

LIWCとは

- ◆ Linguistic Inquiry and Word Countの略称
- ◆ テキスト分析アプリケーション

様々な感情的、認知的、および構造的な要素の研究のために
効率的かつ効果的な方法を提供するために開発

LIWCの遷移

- ◆ 言語研究の調査とその開示の一環(1993)
- ◆ 拡張された辞書とより現代的なソフトウェア設計で元のアプリケーションを更新(01, 07)
- ◆ 辞書とソフトウェアが一新(15)
柔軟さや透過性を意識しユーザの色を出せる

LIWC2015

- ◆ 数字、句読点、さらには短いフレーズにも対応できます

→ "netspeak" language, SNS への対応

LIWCの仕組み(1/2)

- ◆ target wordの決定(どの単語の出現回数を記録するのか)
- ◆ Target words = dictionary wordsの時,
target words所属する category scaleが増加.
- ◆ 単語数, 句読点の数などの構造的要素も記録.

LIWCの仕組み(2/2)

- ◆ 約6,400語, 語幹, select emotionsで構成
- ◆ LIWC2015辞書は階層構造になっている

LIWC作成過程

- ◆ Step1 Word Collection.
- ◆ Step2 Judge Rating Phase.
- ◆ Step3 Base Rate Analyses.
- ◆ Step4 Candidate Word List Generation
- ◆ Step5 Psychometric Evaluation
- ◆ Step6 Refinement Phase
- ◆ Step7 Addition of Summary Variables.

αの作成

Table 2. Summary Information for LIWC2015 Statistics

	Blogs	Expressive writing	Novels	Natural Speech	NY Times	Twitter
Total files	37,295	6,179	875	3,232	34,929	35,269
Total authors	37,295	2,510	441	2,174	Unknown	35,269
Total words	119,449,058	2,526,709	57,467,183	2,566,446	26,007,632	23,172,994

Note: All texts for all corpora required a minimum of 25 words for inclusion in our analyses. All texts with fewer than 25 words were omitted for all statistics reported in this document.

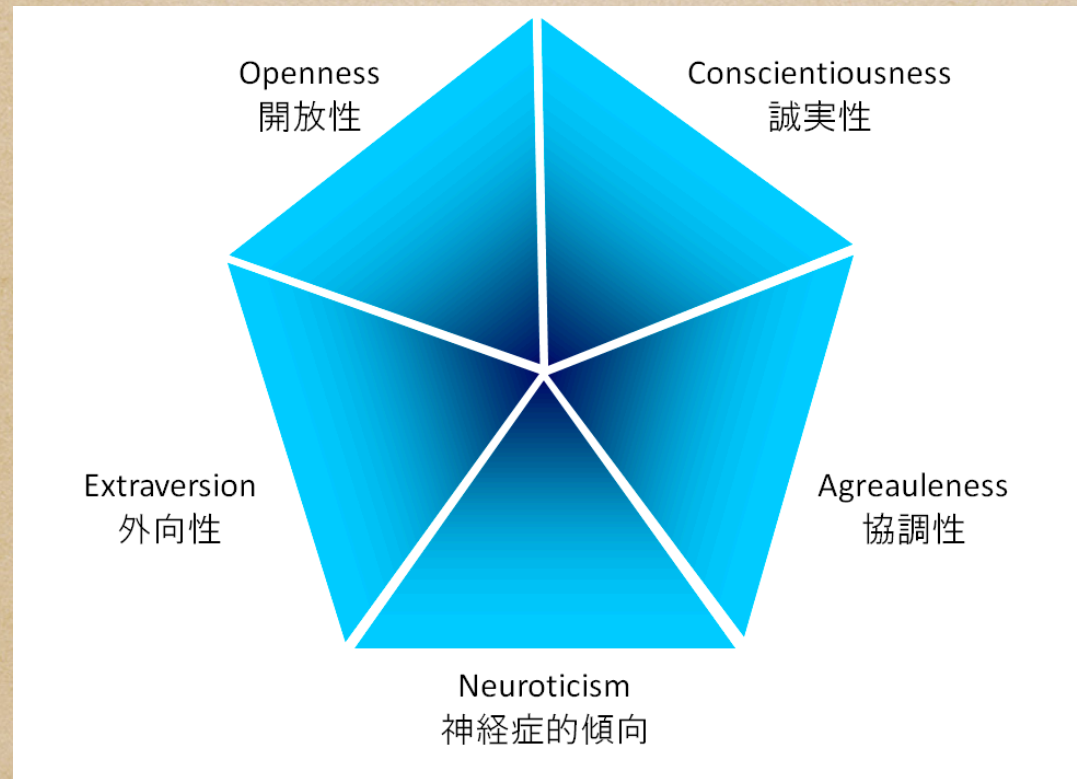
LIWCの信頼性

- ◆ 信頼性を知るにはTausczik and Pennebaker (2010)を読むと良い.

性格分析-BIG5-

- ◆ 外向性~内向性
- ◆ 情緒安定性
- ◆ 開放性
- ◆ 誠実性
- ◆ 協調性

性格特性



◀ 衝動的

良 識 性

秩 序 ▶

◀ 神經質

情緒安定性

樂觀的 ▶

◀ 保守的

知的好奇心

空想的 ▶

◀ 消極的

外 向 性

活 潑 ▶

◀ 個 別

協 調 性

愛 着 ▶

自然言語の信頼性係数の許容範囲

- ◆ 自然言語使用での心理測定は難しい
 - 一度話すとその段落やエッセイでもう一度言う必要がないが会話なら複数回使うから
- ◆ 一般的な心理テストの信頼係数より低くなる