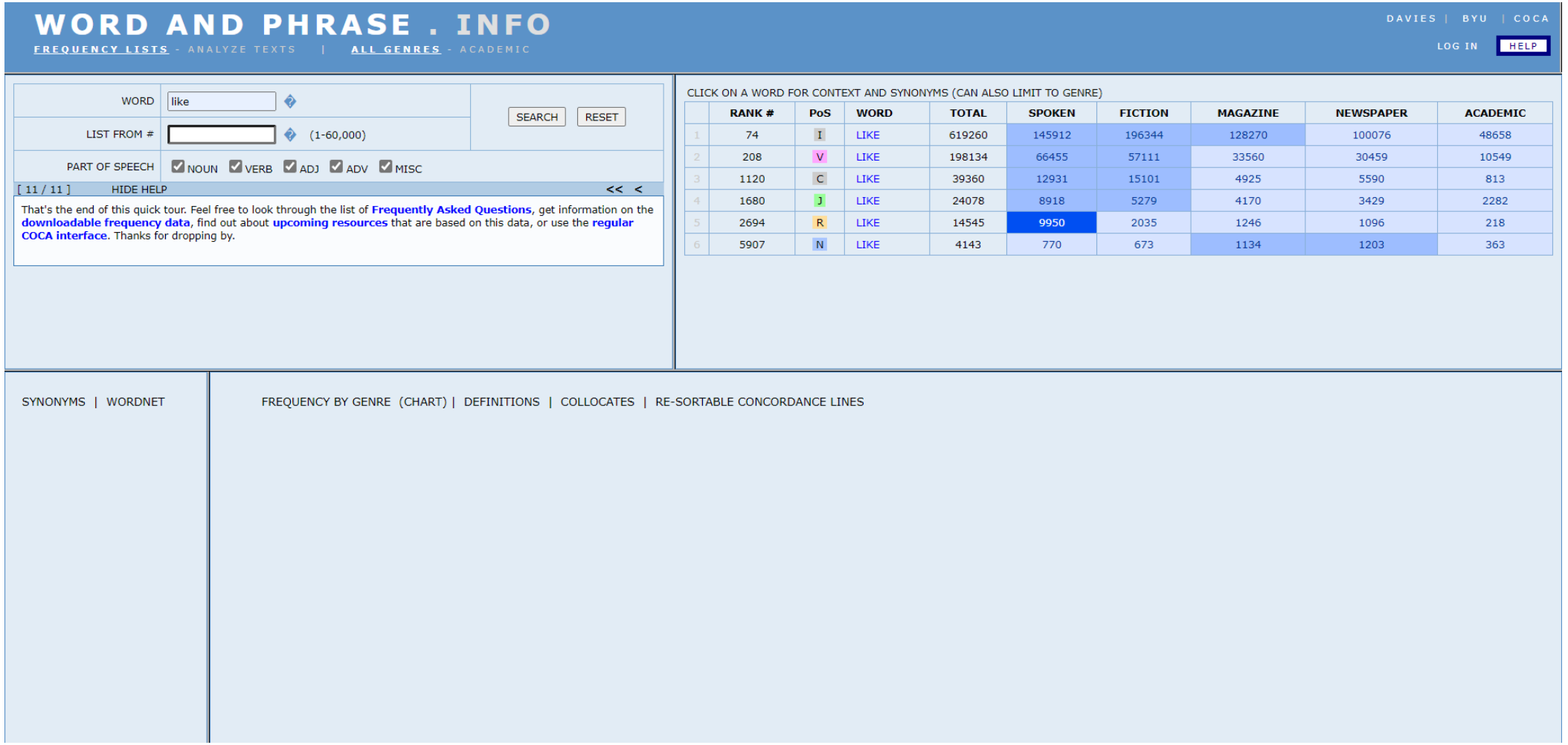
A3.2.3
Other Visualizers and Tools

2020年8月22日

2020年投野ゼミオンライン合宿

ウルドゥー語科3年　8118101　萩原萌夏

☐WordandPhrase.Info



□LexTutor　単語もしくは単語頻度とワードファミリーの関係性示す

□VocabProfile 語彙使用の分析

A3.2.3.1 Sketch Engine

□Sketch Engine

- 近年話題のオンラインのコーパスツールの一つ
- コーパスマネージャー 兼 分析ソフトウェア
- Adam Kilgarriff と Pavel Rychly が開発
- 3つの構成要素：
  ①大規模なデータベースの管理システム(Manatee)
  ②ウェブインターフェイスでの検索プログラム
  ③コーパス作成や管理に役立つウェブインターフェイス

A3.2.3.1　Sketch Engine

- Sketch Engine’s database management system (Manatee)
  Sketch Engineの主要な構成要素の一つ
  大規模コーパスの索引に役立つ

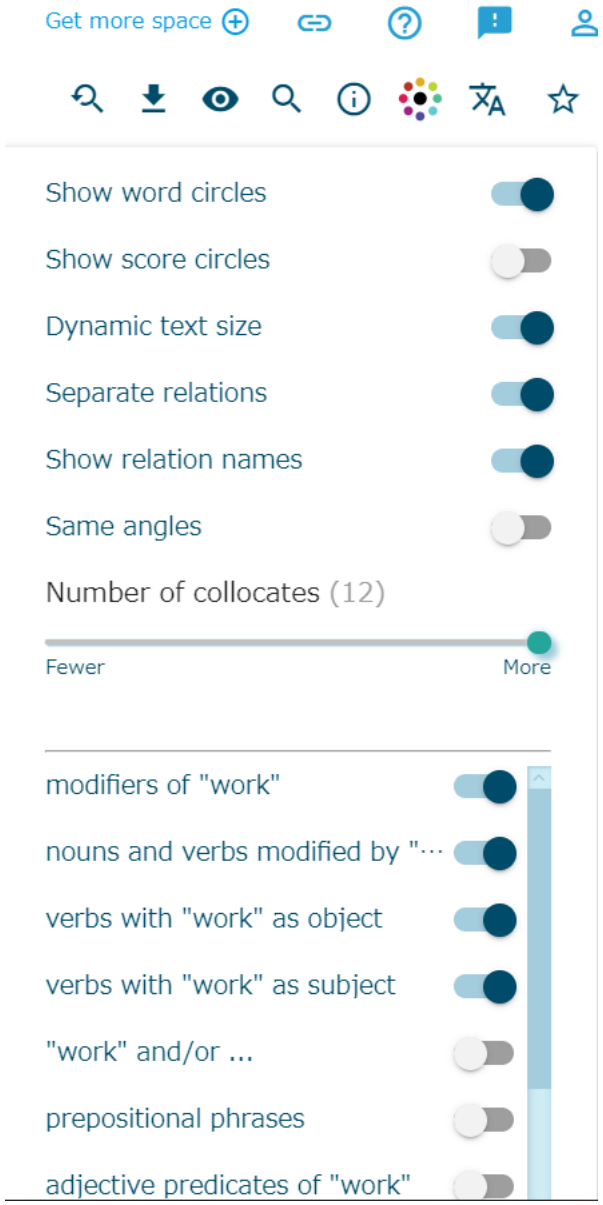
- Corpus Query Language (CQL)
  コーパスから単語や文章のデータを簡単に取り出すことが可能
  対象コーパス数や言語数など年々拡大している
  ※実際に検索してみる

A3.2.3.1　Sketch Engine

- テキストにあるように実際に検索してみた

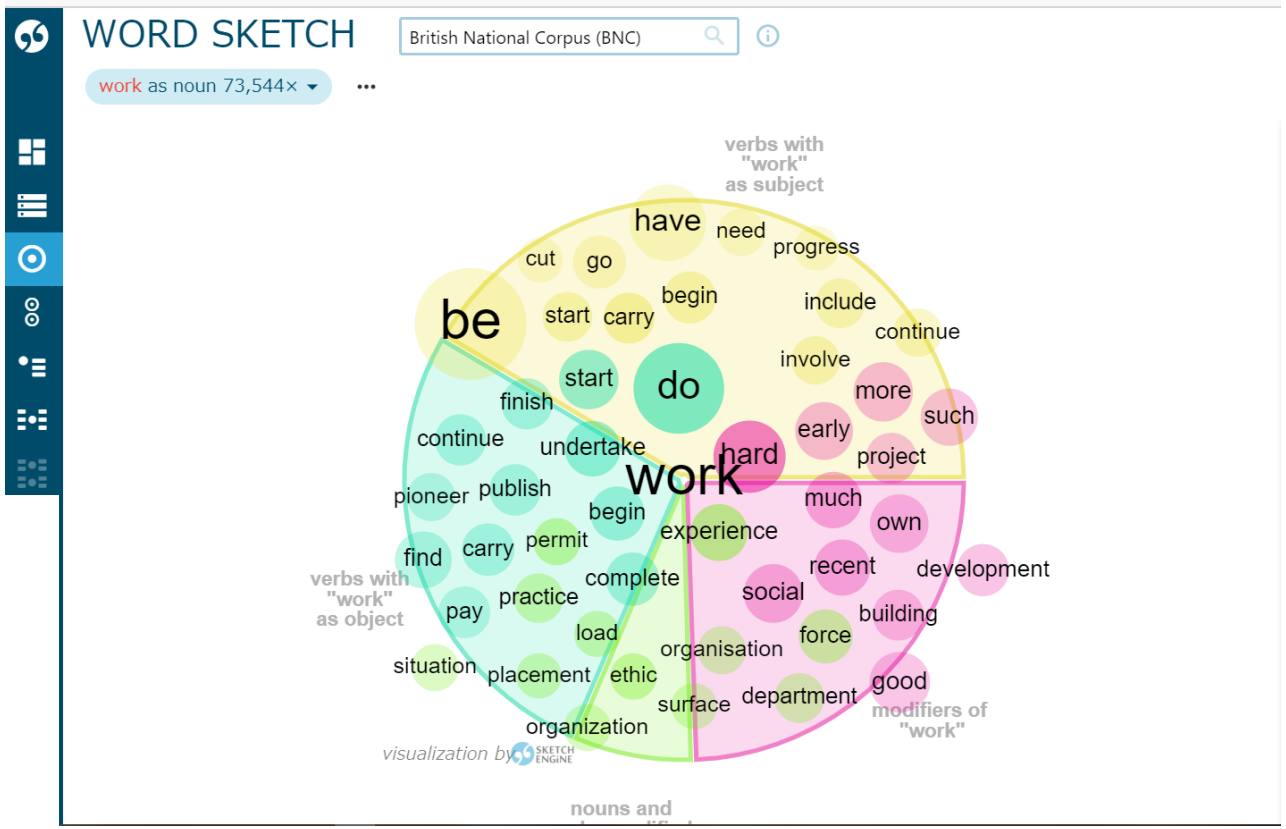
- Sketch EngineでBNCを選択
- Word SketchとThesaurusで名詞“work”を検索
- 右上の“show visualization”を選択

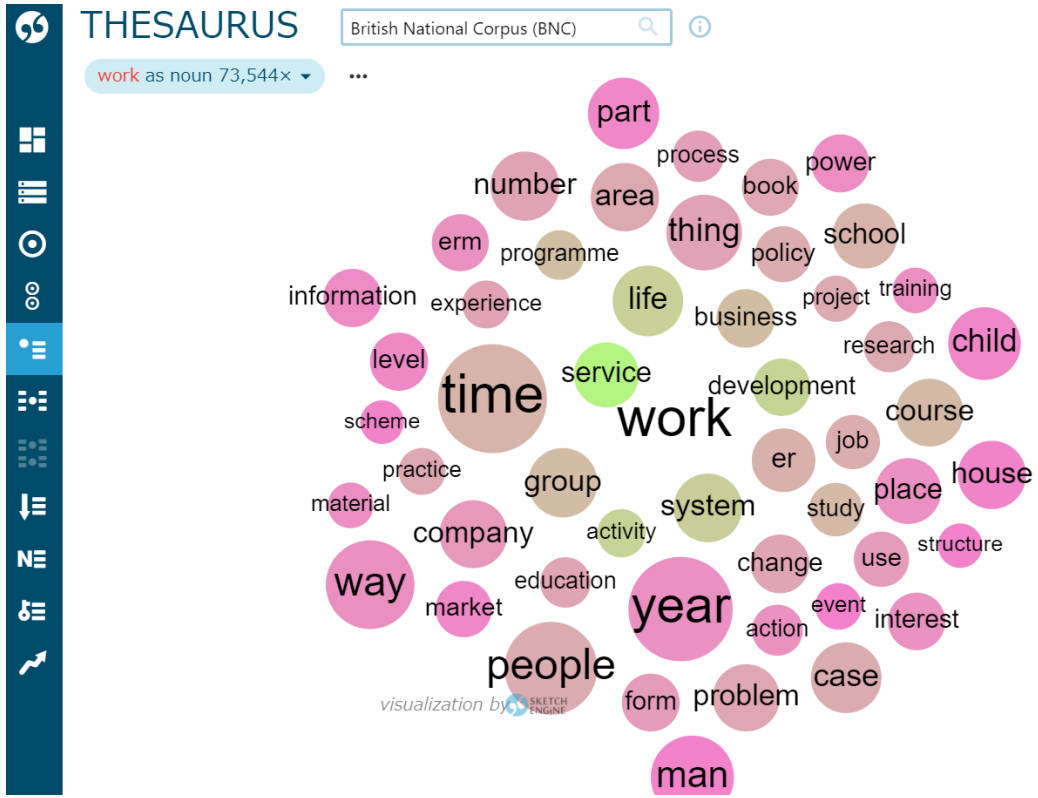
- 結果
- サークル状に単語が表示される
- 様々な表示方法を選択できる
- Word Sketch→共起語、Thesaurus→類義語



A3.2.3.1　Sketch Engine

Word Sketchでの表示→

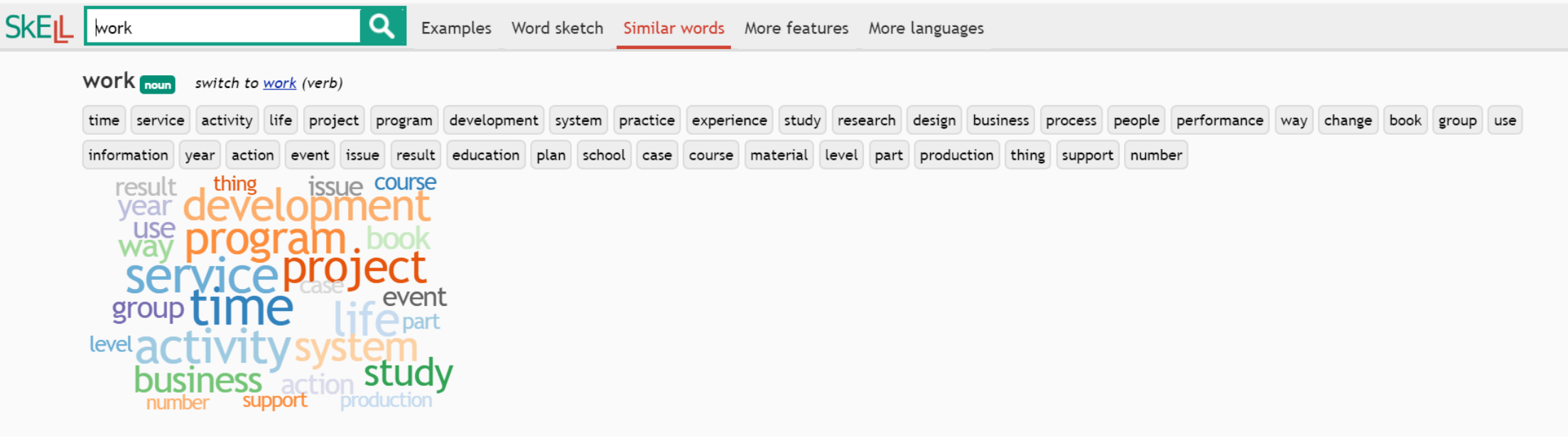




←Thesaurusでの表示

□Sketch Engine for Language Learning (SkELL)

- アップデートver.
- 英語教師、英語学習者向け
- 英語母語話者が実際に使用する単語や文章をコーパスの中で確認することができる



A3.2.3.2
Visualizing Online Language

社会言語学の分野における方言研究

- コーパスを基盤とした方言の変動分析に有効なデータを使用

Grieve(2016)

- 地域別の米英語、米の70以上の都市から収集したコーパス
- コーパス基盤の教育法と視覚化技術が、書き言葉の地域変動に関する教育や調査にどのくらい使用されているかを明らかにする
- アメリカの12の方言地域を表記した新たな方言マップを作成

SNSにおける言語：Twitter

- オンライン言語の視覚化が急激に取り上げられるように
- ソーシャルメディア（Facebook/Twitter など）における会話データ
- TwitterのAPIによってリアルタイムでの情報収集が可能

  例）トランプ大統領のTwitter投稿

  →言語分析＆ビジネス＆大規模データ分析に役立つ
- 基本的に全ての人に公開されるデータ

Eisenstein(2010)

- Twitterの位置情報を基にした調査
- GPSなどの端末情報から地理的座標が特定可能
- ほとんどの投稿にはジオタグが付いている→場所の特定可能

※ジオタグ（geotag）…投稿に付与できる地図上の位置を示す数値データ

- トレンドや使用者情報に着目した研究も

　　→マーケティングやユーザーの行動追跡データ収集に役立つ

Grieve(2014)　Twitterにおける語彙の広がりの調査

- 数十億語の地理的モニターコーパスを使用
  （2013年以降の米英のジオタグが付いた投稿、約2500万語/日）
- 新単語の発生と拡大の過程を特定する

- “**rising** Twitter word”と”**falling** Twitter word”のリスト化
  2013年1月～9月の250日間で最低で1000回登場した6万語を抽出し、使用が増加した語と減少した語を特定した

“rising Twitter word”と”falling Twitter word”のリスト化

Rising Twitter word	
rn	Right now
selfie/s	A photo of oneself
tbh	to be honest
literaly	
bc	because
ily	I love you
bae	babe,baby
schleep	sleep
sweg	swag
yasss	yes

Falling Twitter word	
wat	what
nf	now following
swerve	
shrugs	
dnt	don’t
wen	when
rite	right
yu	you
wats	what’s
yeahh	yeah

A3.2.3.2　Visualizing Online Language

- 単語ごとのデータの視覚化によって、アメリカにおける単語の広がりの追跡が可能に　例）selfie/s
- ヒートマップが特定の単語が含まれた投稿をしたユーザーの場所を示す
- 最初に使用したユーザーの投稿を視覚化
- 性別や年齢などに従って調査できるプロフィール写真もわかる

FOR THE TEACHER

FOR THE TEACHER

- 社会言語学や社会の言語に関するディスカッションを行う

Grieve(2014)　Twitterの語彙に関するパターンの発見

①上昇語彙はグラフにするとSカーブを描く

②頭字語は上昇傾向に、作られた綴りは減少傾向にある

③アメリカの南部と北部でそれぞれ拡大パターンが存在する

④南部の若い黒人女性と、東部と北部の若い白人男性が語彙を生み出している

議論案

- 視覚化されたデータの印象、得られるものは？
- “data/figure”や”excerpts”を説明しなさい
- “word”とは何か？Twitterにおける新出語彙は言語学習においてどんな意味を持つのか？
- なぜTwitterのユーザーは代替用語ではなく特定の単語や文法形式を好むのか説明しなさい