新型コロナウィルス影響下における質問紙作成の試み(1)

予備的検討

黒　沢　学*

A Preliminary study to construct a questionnaire under COVID-19

KUROSAWA Manabu*

Abstract

Many researches in behavioral sciences are restricted under the influences of COVID-19. In this paper, the author tests whether we can construct questionnaires that asks about mental health under COVID-19 with internet corpus such as social networking services. The author reviewed some Japanese researches on COVID-19 and investigated some scales about mental health under COVID-19. Some linguistics characters of the scales such as number of phrases are calculated and tested whether those characters match internet corpus data such as those from Twitter. Some implications on constructing questionnaires are discussed.

キーワード：コーパス言語学，尺度構成，ビッグデータ

Keywords　：corpus linguistics, psychological scaling, big data

１．問題

1．1．　はじめに

昨 2019 年に始まった新型コロナウィルス COVID-19 の感染はいまだに続いている．原稿作成時最新の WHO(世界保健機関)公式発表(2020 年 10 月 27 日集計)によれば，確認感染者数が全世界で 4296 万 6344 人，確認死者数が 115 万 2604 人である．また，我が国でも確認感染者数が 9 万 0710 人，確認死者数 1646 人となっている(厚生労働省発表，10 月 15 日版)．近年あった SARS(重症急性呼吸器症候群)や MERS(中東呼吸器症候群)もコロナウィルス由来であったが，これらと比べても世界的に見ればその影響は桁外れであるといってよい．

この影響が社会生活や経済など広範に及んでいることは周知である．そのうち，ここでは我々大学における研究者について考えてみる．

研究者もこの事態に手を拱いていたわけではない．医学・薬学を中心とした研究者は現在もこの状況を改善するために努力している．学術情報が集まる Elsevier 社の scopus.com を調べれば，COVID-19 をキーワードとする研究が 60172 件ヒットする(10/27 現在)．そのうちの 2/3 以上が医学領域だが(分野 Medicine が 43781 件)，心理学者も 2021 件の研究を残している．しかし，国／地域を日本に限れば，心理学領域での研究はその 1/100 の 23 件に過ぎない．本研究は，このような状況を受けて，少しでも言語心理学の立場から公共の福祉に対して貢献できることがないかを検討するものである．このようなわけで，通常の論文とは異なり試行錯誤的に検討したことがらを残していく形態となっている．

＊未来科学部人間科学系列教授　Professor, Department of Humanities, Social and Health Sciences, School of Science and Technology for Future Life

1．2．　問題

心理学領域からの研究を見ると，やはり一番多いのは精神的健康を主題としたものである．例えば，Fuse-Nagage, Kuroda, Watanabe(2020)はストレスが一定以上の水準にあるかを調べるK6を用いて，茨城大学の新入生のストレスレベルを調べた．その結果として，ストレスが一定以上の学生は先行研究に比べて多いわけではないこと，一方でストレスが高い学生は将来に希望が持てない，遠隔授業について行けないといった不安を抱えていることが示された．

このような精神的健康を測定する研究では，他に不安(Bohlken et al., 2020)を測定する尺度などが用いられる．このように，従属変数として既成の一般的な精神的健康に関する尺度が用いられる一方で，独立変数としての COVID-19 パンデミックに対してどのように接し，感じ，考えているかという尺度は当然ながらまだやや手薄になっている．健康関係の尺度を検索する Health literacy tool shed でも COVID-19 で尺度がヒットしない．そこで，ここでは COVID-19 に対する態度，感情，行動などを測定するための尺度を構成することを試みる．

COVID-19 影響下の問題のもう一つは，従来行われていた行動のうちいくつかが大きく制約されることである．10 月 27 日現在，多くの人が対面するようなことにはかなりの制限がある．音楽／スポーツなどの大規模イベントは徐々に元に戻っているところであるが，本学では従来のような対面授業はいまだにできていない．このような状況で，研究のためにデータを集める従来の方法のいくつかは行うことができなくなっている．

幸いにして，心理学領域では近年多くのデータが web ベースで集められてきた．いくらかの種類のデータについてはいまだに対面に限られているものの，質問紙法などの言語報告については web ベースでのデータ採取について合意が取れているといってよい．そこで，web 経由でデータを収集することにする．

これまでの心理学における尺度構成では，測定したい概念を特定した後，まず関連する文を自由記述を用いて集めるということがなされてきた．この手順は email など非対面によっても行うことは可能であるが，今回のような行動が制約されている環境下での研究という趣旨に照らせば，これも別の手段で行うのが望ましいであろう．

この点について既に研究を蓄積しているのが，大規模コーパスを用いた研究を行うコーパス言語学である．デジタル化された大規模な言語データとその特性を用いた研究が近年この領域では積み重ねられている．

その中で，今回は言語に関するビッグデータである SNS からの記述を取り上げる．それは，先に述べたような自由記述に近いものが日々積み重ねられているのが SNS であると考えられるからである．SNS も各種あるものの，ここで考えるのは特に Twitter である．Twitter はよく知られたウェブを使った短文投稿サイトであるが，Facebook や Line といった他の SNS に比べ，読み手を限らない(Facebook のように発信した情報をクローズドにするかどうかの選択肢が発信者にない)という点で，幅広いデータが得られることが期待される．そこで，Twitter からのデータが COVID-19 に関する質問紙を構成するに足るだけのものになりそうかを検討する．

具体的には，Twitter からの COVID-19 に関する記述を高い精度で集めることができるのか，そしてそれを素材として質問紙にしていけるのかについて，試行錯誤的に作業をしながら検討してみることにする．

2．方法と結果

2．1．　対照データ

まず，できあがりの質問紙について考える．COVID-19 に関する記述を集めるのであれば，～について困っている(行動)，～についてこう思う(態度)，～についてこう感じる(感情)などの記述が並ぶであろう．それらは，従来 X についての態度(・満～感・意識...)尺度と呼ばれてきたであろう．

そのうち，まずここでは～について困っている，という記述を探すことにする．なぜなら，新たな問題が発生したときにはそれへの最初の対応として，援助を必要としている者がどのような援助を必要としているかを把握することは重要になるはずだ

からである.

すると，その質問紙は困難な状況とその困難な状況でどのような援助が欲しいかについて尋ねるようなものとなるであろう. それは心理学ではソーシャルサポートという文脈で研究されてきた. そこで今回は, 先行研究のうちソーシャルサポートを扱った研究を取り上げる.

ソーシャルサポートを扱った研究も非常に多くがあるが，ここでは尺度化されていること，そしてサポートの内容とサポート源を分類していることから, 福岡(2000)の分類による質問紙を取り上げる. 福岡はソーシャルサポートについて 4 つのサポート内容，11 のサポート源に分類してそれらを測定する質問項目を示しており，それが 21 項目にまとめられている．この 21 項目について，文節数と単語数を求め，要約統計量を産出した．文節数・単語数のカウントには Yahoo！Japan の日本語形態素解析 API を利用し，自立語の数を文節数，総単語数から記号類を除いた数を単語数とした. 表を参照.

２．２． データの収集と対比の結果

Twitter のデータは，申請によって API を用いた収集ができる(取れる数などに限定がある)．筆者は今回の目的のために API の利用を申請し認められているのだが, まずここでは研究が行えそうかどうかをみるために, 手動でデータを集めて見ていくことにする.

Twitter をはじめとするは SNS では，内容を整理するために投稿者が自分で#で始まるタグ(ハッシュタグ)をつけることが広く行われている．ここではこのハッシュタグを用いてデータを選ぶことを試みる. ハッシュタグは内容が何に関連しているのかを示す. それは日常のことばで言うと話題であるが，ここではモチーフと呼んでおく.

一方, そのモチーフについて何を述べるかだが，これについては特に投稿上のデファクトルールはないようである. そこでこちらにはハッシュタグをつけずに検索する．つまり，全てのツイートを対象にして，「#コロナウィルス 困った」で検索してみる．すると，結果画面においてはヒットするツイートが画面を下にスクロールするごとに現れるので，ひとまず先頭の 10 ツイートについて，ハッシュタグを除いた部分をテキスト化して利用した. ヒットしたツイートを例示するならば，上から 3 件は(1) 立憲民主党の政策について反応したもの，(2) 副業を指南する本に誘導するもの，(3) 飲食店を応援しようとするもの，であった．これらについても，前項同様に要約統計量を算出した．表を参照.

表: 質問紙／Twitterごとの文節数・単語数の要約統計量

	文節数			単語数	
	質問紙	Twitter		質問紙	Twitter
Mean	10.8	20.6		20.0	43.7
SD	3.4	8.3		7.0	21.3
Skewness	0.2	0.0		0.4	0.0
Kurtosis	-0.5	-1.0		-0.6	-1.0

この表から明らかなように，文節数・単語数ともに質問紙の質問とツイートの間には単純な文節数・文字数で平均値に違いがある．念のために検定すれば，それぞれ $t(29)=4.50, p<.001$，$t(29)=4.46, p<.001$ である.

３．考察

3．1． Twitter を研究に用いることに関して

以上の結果から, この手法で SNS データから質問紙を構成することができるかどうかについて検討する.

まず，文字数と文節数の問題についてである．文字数と文節数は相関が非常に高く($r=.96$)，実質的に同じ指標と考えてよかろう. これがこれだけ違うのは，Twitter には 140 字までの長さの制限はあるものの, 投稿する内容を複数の文にすることについては制限がなく, 複数の文がひとつのツイートとして投稿されることによっていよう．実際，ひとつのツイートの文の数は今回で平均 3.2，これで補正すると 1 文あたりの文節数は 6.4，単語数は 13.7 で，むしろ質問紙の文章より短い．これは，通常質問紙の文はよく練られ誤解の少ないかたちで表記されるという我々の経験とも一致する. 以上のことから，適切に文章を処理することができれば，Twitter の文章から質問項目を構成することは不可能ではないように見える.

ここまでの分析は，どのように Twitter から関連する文章を抜き出してくるか, ということのみに

関わるものであった．言い方を変えると，意味的な処理はほとんどなし(ヒットする用語のみ)で，そこそこ関連のある文章をもってくることが可能であろう，ということである．ただし，問題はむしろここからであって，それらの素材を元にしていかに有用な質問項目のみを残し，不要な記述を除くか，という問題は手つかずで残っている．

3．2．　今後の方向

以上，自然言語処理的手法を施す前段階としての予備的検討の結果として，Twitter 上の書き込みを組織的に収集することは技術的に可能であり，行うことで意義のある研究につながり得ることを示した．後には，実際に大量のデータに対して各種の自然言語処理的な手法を施すことによって必要な情報を残しつつ，不要な情報を落として質問紙を構成するには，何をどうすればよいかという問題が残る．

これは中核的な問題だが，現在のところさまざまな手法が提案されている．例えば，投稿の(2)に見られたような，ハッシュタグで検索されることを意図した意味的に関係のない書き込みは，そこに使われている単語どうしの意味的な関連の強さ(多次元意味空間での類似度)を用いることで排除することができるだろう．また，「困った」ことを分類したいのであれば，KWIC 形式でデータを整理することでどのような困り方があるかを一覧することができるであろう．これらを用いて自由記述から項目選び，予備調査に続く本調査というかたちでなく質問紙が作れるのであれば，それは現在のような状況に資するものになろう．今後，これらの分析を順次行うことで本研究を社会的な意義のあるものに近づけていくことができると考える．

引用文献

Bohlken, J., Schömig, F.b, Seehagen, T.c, Köhler, S.c, Gehring, K.c, Roth-Sackenheim, C.c, Matschinger, H.a, Riedel-Heller, S.G. 2020 Experience of Practice-Based Psychiatrists and Neurologists during the COVID-19 Pandemic. Psychiatrische Praxis, 47(4),pp. 214-217.

福岡欣治 2000 ソーシャルサポート内容およびサポート源の分類について　日本心理学会第 64 回大会発表論文集．144.

Fuse-Nagase, Y., Kuroda, T., Watanabe, J.　2020 Mental health of university freshmen in Japan during the COVID-19 pandemic: Screening with Kessler psychological distress scale (K6).Asian Journal of Psychiatry, 54.

Health literacy tool shed. http://healthliteracy.bu.edu/all (2020 年 10 月 27 日アクセス)