The University of Osaka
Institutional Knowledge Archive

Title	語彙のネットワークについて ： ロシア語の連想連鎖
Author(s)	上原, 順一
Citation	言語文化研究. 2016, 42, p. 25-42
Version Type	VoR
URL	https://doi.org/10.18910/56202
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

語彙のネットワークについて
—ロシア語の連想連鎖—

上　原　順　一

Network of Words: Chainlink of Association in Russian Language

UEHARA Junichi

Summary: The purpose of this study was to explore the properties of word association networks（chainlink）in Russian language. In the first and the second sections of the study I introduced some basic notions of graph theory, comparing an ideal tree diagram with the real association networks. In the third section, I demonstrated analysis of word networks of root words as *быть*“to be”, *рука*“a hand”and *хороший*“good”. These networks have 5 or 8 thousands of words（as vertices）and 8 or 16 thousands of association relations（as edges）. In this study the main stress falls on the explanation of these big networks. I concluded that they have two main properties: small-worldness and scale-freeness.

キーワード：語彙,ロシア語,ネットワーク

研究の発端

語と他の語とのつながりをここでは語彙のネットワークと呼ぶことにする。このネットワークには実にさまざまな種類がある。まず,実際に発話された文における語のネットワークがある。いわゆる統語的な関係である。これに対して,範列的な関係としては,語彙体系における同義関係,反対関係などがある。いずれも一般的には多くの語からなるネットワークになる。これに対して,通常は 2 語の関係で研究されてきた対象がある。それは連想関係のネットワークである。これは,任意の語（刺激語）を話者に提示して,それから連想する語句（反応語）を答えてもらい,両者の関係を何らかの基準で求めるものである。たとえば,英語の単語連想実験において抽出された重要な反応語タイプとして「等位語,共起語,上位語,同義語」があることが知られている [エイチソン 2010:113-114]。ロシア語では総合的な資料として『ロシア語連想辞典』[Караулов ら 1994] などが知られている。また,これまでに出たロシア語の連想辞典を横断的に分析して,連想語と刺激語の対応関係の差異について調べた研究もある。論文 [Уфимцева

ら2004]など。しかし,刺激語から得られた反応語を新たな刺激語として得られる,いわば,連想の連鎖に焦点をあてたものは少ないように思われる。連想の連鎖とは,「川」->「水」->「透明」などの連想ゲームのようなものである。拙論では,ロシア語におけるこの連想連鎖をネットワークとみなし,この性質をロシア語の高頻度語を素材として調べることにした。この問題への関心は,まず,筆者の知る限り,連想連鎖がロシア語学では未開拓の領域であることに由来する。また,資料となる連想辞典が刺激語と反応語からなるペアの記述であることから,ペアをとりまとめた結果である連想連鎖が複雑かつ大規模なものであり,その分析には対応するソフトウェアが必要になるが,R と Cytoscape を用いて,個人のレベルでも研究が可能になってきたこと,これも調査のきっかけになった。さらには,ネットワークの研究が進むとともに,初学者にもわかりやすい解説書が多数登場し,語と語の連想関係は,語を頂点,連想関係を辺と考えれば,連想連鎖の性質がグラフ理論をもとに記述可能であることが筆者に見えてきたことも,調査の後押しとなった。このように,研究の関心は,領域の未開拓性と可能性である。次のセクションでは,理論的に得られる樹形図と実際に得られた連想関係の縮図をもとに,理論的な説明をしたい。

樹形図と連想関係

まず,分析に利用した資料について簡単に述べる（詳細は [手順番号: 1]）。連想関係を求めるには,最低 1 つの刺激語が必要である。この刺激語として,筆者は主要品詞の中で頻度の大きい быть「ある･いる」, рука「手」, хороший「良い」を選んだ。ここでは,быть を例にあげる。オンラインの「ロシア語連想辞典」（Русский ассоциативный словарь）[PAC] に быть を入力し,反応語などの結果を得る。得られた反応語 201語をさらに刺激語として入力し,結果を得る。この201語には,たとえばчеловеком「人（造格）」がある。これから,быть, стать「…になる」などの反応語が返される。この作業を繰り返し行う。このようにして得たグラフをそれぞれ,「быть グラフ」,「рука グラフ」,「хороший グラフ」となづける。быть, рука, хороший から得られる反応語のうち刺激語を返さない語は,反応語としては扱わない。

利用した「ロシア語連想辞典」は,100万を超える連想関係,刺激語は6000以上,反応語は 10万以上あるので,データの数は十分であると考えられる。また,この「連想辞典」はこれまでに出版された各種連想辞典を基盤にしており,信頼性もあると思われる [PAC]。

グラフ理論について簡単に述べる。グラフとは「いくつかの点と,それらの点を結ぶ線からできている図のことである」。グラフの例としては,鉄道網がある。ここでは駅が点,線路が線である [小林 2013: 2]。この「点」と「線」には,さまざまな用語が当てられているが,拙論ではそれぞれ「頂点」,「辺」と呼ぶ。語と他の語の関係は,グラフ理論では,語が頂点,関係が辺であるとみなすことができる。ただし,その関係には向きがある。語 A から連想される語 B がある場合,これは A から B であって,その逆ではない。そのため,辺には向きをもうける必要が

ある。辺に向きがあるグラフを有向グラフと言う。

筆者が得た資料を用いると,刺激語と反応語のペアだけではなく,さらに多くの語を巻き込んだ,いわば連想の連鎖を研究対象にすることが可能である。たとえば,次の図を参照してもらいたい。

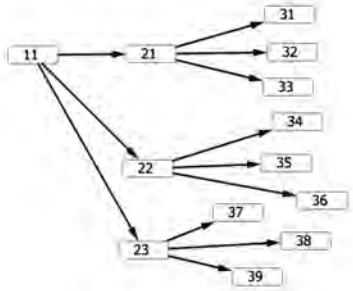

図 1　樹形図の例

これは,連想辞典で根としての刺激語（11）が 他の語（21, 22, 23）への連想関係をもっているだけではなく,連想の反応語をさらに連想辞典で見ると,その反応語（たとえば 21）が刺激語となり,それと反応語（31, 32, 33）との間に連想関係があることを図示したものである。もちろん,このような連想の連鎖を得る場合でも,任意の刺激語をひとつ設定して,これからの連鎖を求めることになる。この連鎖はひとつの樹形図のような形式をとる。樹形図のはじめに設定する語（図 1 の語11）はグラフでは根と呼ばれる。

この樹形図は,連鎖のもっともわかりやすい形ではあるが,実際の連想連鎖はやや複雑である。たとえば,連想辞典では рука が刺激語である連想関係として,“рука -> нога”,“рука -> друга”,“рука -> правая”が得られる。これらの反応語を刺激語として設定すると,“нога -> рука”,“нога -> болит”,“нога -> больная”,“правая -> рука”,“правая -> левая”,“правая -> нога”,“левая -> нога”,“левая -> рука”,“левая -> сторона”の関係も得られる。この連鎖関係を例示するために,рука から得られた反応語のうち最も反応件数が多い 4 語,нога, друга, правая, левая とそれらを刺激語として得られた 3 語 рука, болит, больная, рука, левая, нога, нога, рука, сторона から成り立つグラフを「ruka モデル」と名付ける [手順番号: 2]。これは,連想の連鎖を表す,小さな縮図・模型のようなものである。

まず,この刺激語（S）と 反応語（R）の関係と語の日本語訳を表に示す。

表１　ruka モデルに出る語の日本語訳

S と日本語訳		R と日本語訳	
рука	手	нога	足
рука	手	друга	友人の
рука	手	правая	右の
нога	足	рука	手
нога	足	болит	痛む
нога	足	больная	病んでいる
правая	右の	рука	手
правая	右の	левая	左の
правая	右の	нога	足
левая	左の	нога	足
левая	左の	рука	手
левая	左の	сторона	側

語の連想関係を図示すると下のようになる。

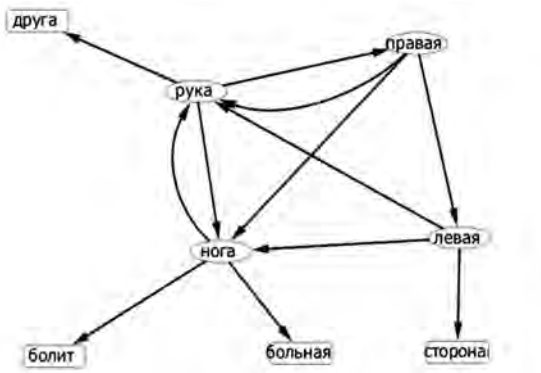

図２　ruka モデル

規則的な樹形図と実際に得た ruka モデルを比べることによって,グラフにおける連想関係の特徴を探ることにする。

グラフでは,頂点に入る辺の数を入次数,頂点から出る辺の数を出次数と呼ぶ。これらの合計,すなわち,頂点に接している辺の数を次数と呼ぶ。まず,これを基準に比較する。

まず,入次数について。樹形図では,入次数 0 の頂点は根のみである。入次数 1 の頂点は根以外の頂点である。一方, ruka モデルでは入次数 1 の頂点は правая「右の」など 6 個,入次数 2 の頂点はないが,入次数 3 の頂点は рука「手」以外に нога「足」である。ruka モデルは,根を欠くものに性質が変異している。これは図 1 の樹形図では語11にあたる根が図 2 の ruka モデルにはないということである。本来的に樹形図の根は入次数が 0 であることが特徴的である。ruka モデルは, рука「手」をはじめに設定したグラフであるが,この語を反応語とする連想関係があるためにこの入次数が 0 ではなくなり,かつ,他に入次数が 0 である頂点が他にないので,根を欠くと言えるのである。

表２　入次数

入次数	0	1	2	3
樹形図	1	12	0	0
ruka モデル	0	6	0	2

次に,出次数について述べる。樹形図では出次数 0 の頂点は末端の頂点に限られ,この頂点数は 9 である。出次数 1 と出次数 2 の頂点はなく,出次数 3 の頂点数が 4 である。この出次数 3 の頂点は,樹形図の根と末端頂点以外の頂点である。これらは,いわば,樹形図の途中段階で出現する。一方, ruka モデルでは,出次数 0 の頂点数が 4 である。これは連想連鎖の末端である。出次数 1 と出次数 2 の頂点はない。特徴的なのは,最大の出次数 3 をもつ頂点がрука「手」以外にもあることである（нога「足」, правая「右の」, левая「左の」)。本来, 図 2 の ruka モデルはрука「手」から始まる連想関係なので,この語が刺激語として重要な意味をもち,大きな出次数を有することは容易に想定できる。しかし,この語と同じ最大の出次数 3 をもつ語が複数あることは,このモデルを作成してはじめてわかることである。一方,図 1 の樹形図では,この次数 3 をもつ頂点は, 先述のように根と末端以外の頂点なので,樹形図の性質から求めることが可能である。

表３　出次数

出次数	0	1	2	3
樹形図	9	0	0	4
ruka モデル	4	0	0	4

頂点の入次数と出次数の合計を次数と呼ぶ。一般に,次数の大きい頂点をハブと呼ぶことがある。これは,情報や交通の流れを見た場合,中心的な役割を果たす頂点のことである。樹形図では次数は 4 が最大で,この次数をもつ頂点は樹形図の中間にある頂点である。樹形図の根から他の頂点に情報が流れることを想定すると,この中間頂点の役割は重要である。グラフ理論では,頂点の重要性をみる基準のひとつに次数中心性がある。これは,次数の大きな頂点がグラフにおける中心,いわゆるハブのようなものとみなす考え方である。これは連想連鎖においても同様であると考えられる。ruka モデルでは,次数の最大は 6 であり,この次数をもつ頂点はрука「手」と нога「足」である。これは,рука「手」からはじまる連想の体系で,この語以外にнога「足」も他の語とのつながりという点で重要な役割を果たすことが示唆されている [表 4]。

表４　次数

次数	0	1	2	3	4	5	6
樹形図	0	9	0	1	3	0	0
ruka モデル	0	4	0	0	2	0	2

グラフ理論で提案されている中心性のひとつは,上述のように次数中心性である。中心性にはあとひとつ媒介中心性がある。これは,他の複数の頂点との結びつき具合をもとにその頂点の重要性を測る指標である。これは,「そのノード（頂点）を通過しないと他のノード（頂点）に到達できない度合い,ある点がその他の 2 点を結ぶ最短経路である度合いであり,値が大きいほど中心性が高い」[金 2009:65-67]。

語の連想連鎖では,語から他の語への連鎖において,別の語が経由していることが多くみられる。この,いわば経由地となる語は,やはり連鎖において重要であると考えられる。これを求めてみよう。頂点の媒介中心性は,その頂点を通る経路数の総和をその頂点を除く頂点の組み合わせ数で割ったものであると定義されている。

これを ruka モデル（図 2）でみてみよう。рука を通る経路は,“нога「足」-> рука「手」-> правая「右の」”,“нога「足」-> рука「手」-> правая「右の」-> левая「左の」”,“нога「足」-> рука「手」-> друга「友人の」”,“нога「足」-> рука「手」-> правая「右の」-> левая「左の」-> сторона「側」”,“правая「右の」-> рука「手」-> друга「友人の」”,“левая「左の」-> рука「手」-> правая「右の」”,“левая「左の」-> рука「手」-> друга「友人の」”の 7 個である。рука「手」以外の頂点数は 7 であり,これから 2 つ選ぶ組み合わせ数は 42 である。従って,рука「手」の媒介中心性は約0.17 となる [手順番号: 3]。

ruka モデルで рука「手」と同じく次数中心性が高い нога「足」をみよう。нога を通る経路は,“рука「手」-> нога「足」-> болит「痛む」”,“рука「手」-> нога「足」-> больная「病んでいる」”,“правая「右の」-> нога「足」-> болит「痛む」”,“правая「右の」-> нога「足」-> больная「病んでいる」”,“левая「左の」-> нога「足」-> болит「痛む」”,“левая「左の」-> нога「足」-> больная「病んでいる」”の 6 であるから,媒介中心性は約0.14である。つまり,この 2 語は,グラフ理論で言及されている中心性をみる場合,次数中心性と媒介中心性は異なり,経由地な役割においては рука「手」が優位にあることがわかる。

次のセクションでは,以上でとりあげた概念を中心に,быть, рука, хороший の 3 グラフを観察する。

３つのグラフ

まず,быть, рука, хороший それぞれのグラフで媒介中心性の大きな 20 語を抽出してみよう [手順番号: 4]。

表５　媒介中心性の大きい20語

グラフ	быть	рука	хороший
1	быть	рука	хороший
2	хорошо	человек	большой

3	жизнь	голова	плохой
4	человек	работа	человек
5	здесь	помощь	прекрасный
6	жить	друг	праздник
7	я	сердце	я
8	вместе	моя	очень
9	есть	нос	друг
10	друг	жена	добрый
11	любить	палец	рабочий
12	находиться	сила	товарищ
13	он	кулак	день
14	иметь	делать	вечер
15	видеть	река	ответ
16	стоять	нога	лес
17	умереть	показывать	парень
18	год	отец	солнце
19	являться	товарищ	желтый
20	тащить	тело	он

それぞれのグラフで媒介中心性の大きな語には根である語と似た形態的ないしは意味的な特徴がみられる。быть グラフでは,動詞 быть「ある・いる」と同じ品詞である語が多い(быть, жить「住んでいる」, любить「愛している」など 11語)。рука グラフでは,рука「手」と同様に人体の一部を表す語が多い(голова「頭」, сердце「心臓」, нос「鼻」など 8 語)。хороший グラフでは,形容詞が多い(хороший「良い」, большой「大きい」, плохой「悪い」など 6 語)。また,統語的につながりうる語もみられる。( ) 内は語の変化形を示す。それぞれ, “быть「ある・いる」- хорошо「よく」, здесь「ここで」, вместе「一緒に」”,рука「手」- человека (человека)「人の」, друг (друга)「友人の」, моя「私の」, жена (жены)「妻の」, отец (отца)「父の」, товарищ (товарища)「同僚の」, рука (рукой)「手で」, “хороший「良い」- человек「人」, праздник「祝日」, друг「友人」, день「日」, вечер「夕方・パーティー」, ответ「答え」, лес「森」, парень「青年」” など 。当然,根となる語と直接的な連想関係にある語が多いので,記載した語の媒介中心性が大きいのはかなりの程度予測できる。ただし,そもそも媒介中心性の大きな語は,根である語との関係のみで存在するわけではない。たとえば,表 5 における человек「人」である。これを経由地とする連想の連鎖は, быть グラフでは, “я「私」-> человек「人」-> счастливый「幸せな」”, “смелый「勇敢な」-> человек「人」-> сердечный「心のやさしい」” などがみられる。なお,この человек「人」が中間にある経路数は 37000 を超えることがわかった。同様に,рука グラフ における друг は “товарищ「同僚」- друг「友人」- большой「大きい」”, “человек「人」- друг「友人」- собака「犬」” などの連鎖の経由地で,この語のこのような連鎖経路は 23000 を超える。хороший グラフにおける плохой「悪い」は “такой「そのような・とても」- хороший「良い」- юмор「ユーモア」”, “худший「より悪い」- плохой「悪い」- член「メンバー」” などで,連想連鎖は 10900 を超え

る[手順番号: 5]。

一方で,複数のグラフに共通して登場する語もある。それは, человек「人」, друг「友人」, я「私」である。品詞も語彙的意味も異なる語を根とするグラフでこの3語の媒介中心性が高いと言うことは,拙論ではとりあげなかった連想連鎖をグラフにしても,これらの語は媒介中心性が大きい,つまり連鎖の橋渡し的な役割において重要な語ではないかと推測できる。

次に,連想ネットワークが意外と小さいことについて述べる。「意外と」というのはグラフの頂点数,辺数と比べた場合のことである。3つの語 быть, рука, хороший のグラフは,頂点つまり語の数がそれぞれ 5452, 5383, 8874 である。頂点と頂点,つまり語の連鎖を表す辺の数はそれぞれ 8379, 8030, 16715 である[手順番号: 2]。連想ネットワークがこの大きさに比べて意外と小さいのは,語からたとえ他の語を経由するにしても別の語にたどりつく連鎖のステップである。頂点から別の頂点への辺の数,つまり連鎖のステップは小さく,任意の語から別の語への平均的なステップ数(距離)は,おおむね3である。この点について,まず, ruka モデル[図2]について確認してみよう。

ruka モデルの語から別の語への最短経路は,次のようになる。数字は距離である[手順番号: 6]。

表6 ruka モデルの最短経路

рука から他の語への経路	
“рука -> нога”	1
“рука -> правая”	1
“рука -> правая -> левая”	2
“рука -> друга”	1
“рука -> нога -> болит”	2
“рука -> нога -> больная”	2
“рука -> правая -> левая -> сторона”	3
нога から他の語への経路	
“нога -> рука”	1
“нога -> рука -> правая”	2
“нога -> рука -> правая -> левая”	3
“нога -> рука -> друга”	2
“нога -> болит”	1
“нога -> больная”	1
“нога -> рука -> правая -> левая -> сторона”	4
правая から他の語への経路	
“правая -> рука”	1
“правая -> нога”	1
“правая -> левая”	1
“правая -> рука -> друга”	2
“правая -> нога -> болит”	2
“правая -> нога -> больная”	2
“правая -> левая -> сторона”	2
левая から他の語への経路	
“левая -> рука”	1
“левая -> нога”	1

“левая -> рука -> правая”	2
“левая -> рука -> друга”	2
“левая -> нога -> болит”	2
“левая -> нога -> больная”	2
“левая -> сторона”	1

これらの経路の数は 28 で,距離の合計は 48 である。経路の平均距離は約 1.7 である。これは,語から別の語への連想連鎖がある場合,平均すると 1.7 のステップで到達できることを意味する。これは, [手順番号: 7] でも求められる。

これを 3 語のグラフで求めると,それぞれ,3.1, 3.4, 3.2 となった [手順番号: 8]。

グラフにおける経路の短さは比喩的に「スモールワールド」と呼ばれている。これは,様々な実験により,経験的に知られている。有名な知見を記す。まず,ハーバード大学教授スタンレー・ミルグラムの実験である。この実験では,互いに知らない人どうし X から Y まで,知り合いを介して手紙を送り続けると,平均して 6 通の手紙で到着するとの結果が得られた。似たような実験が コロンビア大学のダンカン・ワッツによって行われた。これは手紙ではなくメールによるものであるが,スタートとゴールの人が同じ国ならば 5 通で,違う国ならば 7 通で到着したというように,結果としてはミルグラムの場合と似ている。比較的少ない隔たりで点と点がつながっていることは,ウェブサイトのリンク,会社の持ち株関係,電力網など,あらゆるネットワークのもつ性質だと考えられている [増田・今野 2006: 60-66]。拙稿で試みた быть, рука, хороший のグラフでも語数が多い割には任意の 2 語がおおむね 3 というステップでつながっているので,一般的なネットワークの性質を備えていると考えられる。

次に,頂点がもつ次数を調べてみる。それぞれのグラフで次数の大きい20語は次の通りである。数字は次数である [手順番号: 2]。

表７　３グラフで次数の大きい語

быть グラフ	рука グラフ	хороший グラフ
человек,308	рука,404	человек,351
я,274	работа,320	я,272
находиться,273	человек,311	большой,256
жизнь,253	показывать,228	плохой,255
есть,248	гнев,226	хороший,255
умереть,248	сердце,219	начальник,250
являться,247	отец,219	праздник,248
любить,231	брать,210	результат,239
хорошо,230	голова,209	песня,232
ждать,230	помощь,209	разговор,231
быть,228	жена,208	отец,231
видеть,226	делать,201	урок,230
существовать,226	нога,200	они,226
здесь,224	товарищ,189	парень,225

жить,222	просто,188	муж,225
стоять,221	нос,183	вор,223
год,218	давать,179	голос,220
тащить,215	река,177	мало,220
иметь,210	дать,172	город,219
остаться,209	палец,161	спектакль,217

これをみると,グラフによっては,当初の刺激語が必ずしも最大の次数をもつわけではないことがわかる。быть グラフと хороший グラフで最大の次数はいずれも человек「人」である。また, человек「人」, я「私」は複数のグラフで高い次数を有している。これは,他の語が連想の根とするグラフを作成しても,共通して出現する次数の高い語,つまり,連想の連鎖でハブの役割を果たす中心的な語が存在する可能性を示唆していると考えられる。

この次数については,その分布がグラフ理論でよく用いられる。分布は,次数ごとにその次数をもつ語数を求めることで得られる。拙稿でしらべた 3 グラフの次数分布をみてみよう。[手順番号: 9]

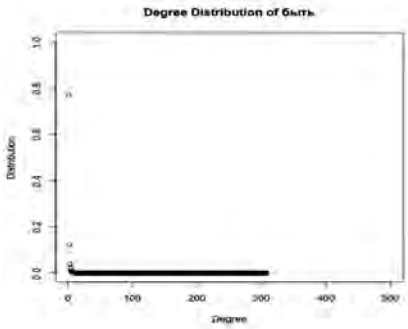

図 3　быть グラフにおける次数分布

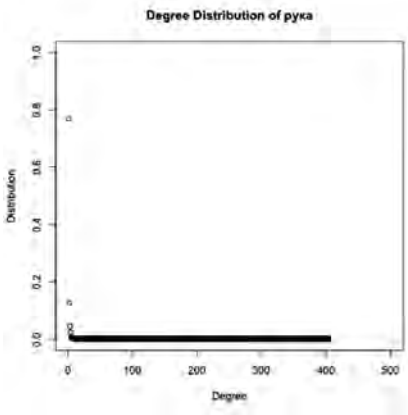

図 4　рука グラフにおける次数分布

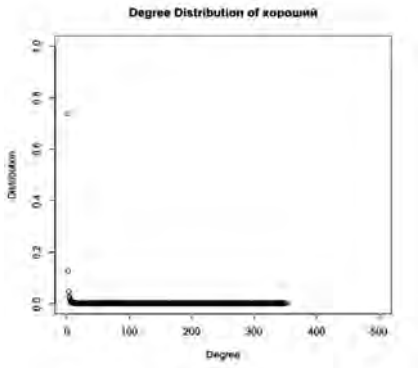

図5　хороший グラフにおける次数分布

それぞれ,横軸が語の次数で,縦軸がその次数をもつ語の分布割合（1.0 が100パーセント）である。一見して,次数の低い語（つまり他の語との連想関係が少ない語）が非常に多い一方で,次数の高い語（つまり他の語との多くの連想関係でつながっている語）が非常に少ないことがわかる。ただ,特に次数の非常に大きい,または小さい語のあたりの分布が上の図では見にくいので,両対数グラフにしてみよう [手順番号: 10]。

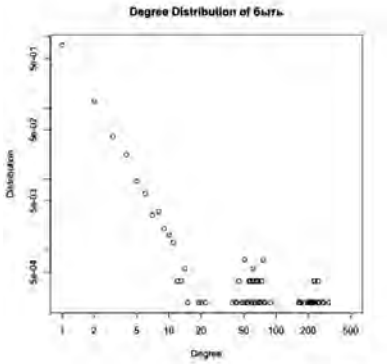

図６　быть グラフにおける次数分布（両対数グラフ）

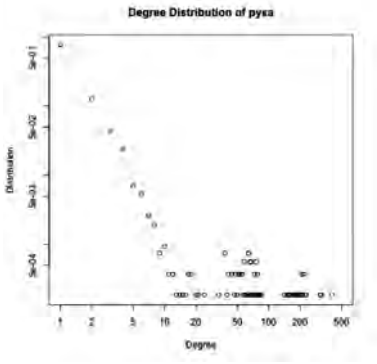

図７　рука グラフにおける次数分布（両対数グラフ）

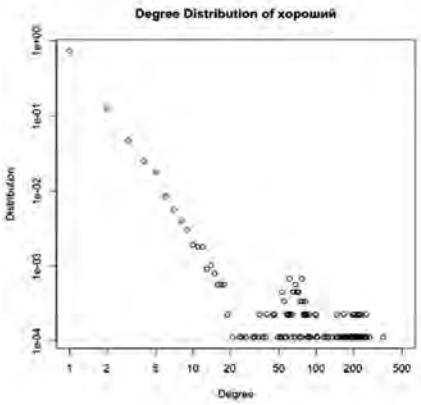

図８　хороший グラフにおける次数分布（両対数グラフ）

これらのグラフより, 次数がおおむね 1 から 20 あたりの小さい範囲では,分布は左上から右下にさがる直線であることが推定できる。両対数グラフにおけるこの直線は,一般に y = x^a（^a は累乗またはべき乗）を示している。このように分布がべき乗になることは,多くのネットワークでも同様であることが知られている。たとえば,頂点がウェブサイトであるウェブのネットワーク,頂点が俳優である映画俳優の共演ネットワーク,頂点が空港の航空網なども,次数分布がべき分布をなすという [増田･今野 2006: 103]。ネットワークのべき則は,「スケールフリー性」があると言われている。特徴的なスケールとは,たとえば,つりがね状分布における平均値や分散のことである。べき分布のグラフでは,これらの性質が欠けているのでスケールフリーと考えられているのである。拙稿でとりあげた быть, рука, хороший のグラフでは部分的にではあるが,これらのネットワークの特徴を示している。

「スモールワールド」や「べき分布」,「スケールフリー」については, Barabasi A.L[2002],Caldarelli G. ら [2012] など,ネットワークの研究で多く言及されている。

まとめ

言語学者ガーク（Гак）は,青少年向けの言語学事典で次のように述べている。「言語における語は孤立して生きているのではない,語は異なった大きさをもったグループ（体系）にまとまっている」[Гак 1984: 144] と。言うまでもなくこれはメタファーであり,「語は人々と同じく孤立して生きているのではない,語は人々と同じくグループにまとまって生きている」とのように解釈できる。

グラフ理論が応用されている様々な分野では,たとえば,鉄道網のネットワークでは駅が頂点と,線路が辺とみなされている。同じように,人間のネットワークでは人が頂点のように,人付

き合いが辺と考えられる。この何かを頂点に,何かを辺に読み替える作業がネットワーク科学の発想に新しい視点を持ち込んでいるように思える。つまり,頂点や辺とみなされるものは,他にもあるのではないかという発想である。これもメタファーである。

拙論は語を頂点と,語の連想関係を辺とみなし,これらの総体をグラフと考えることからスタートした。このメタファーをもとに作成した,連想関係の連鎖からなるグラフは,語が非常に多いにもかかわらず,語から他の語への連鎖は意外と短いことを明らかにした（スモールワールド)。そして,他の語と非常に多くのつながりをもつ語は少なく,大多数の語が他の語とのつながりを少ししかもたないことをみた（スケールフリー)。これがメタファーから得た結論である。

言語をメタファーとして語ることにはどのような意味があるだろうか。これについて,言語学者 ラヒーリナ（Рахилина）は述べている：「科学分野におけるメタファーはまるで『反作用のように機能して』,この分野における観念的基盤をすべて再構築することがたびたびある」と [Рахилина 2000: 11]。語の関係をグラフというメタファーをもとに調べていくと,この研究分野に新たな理論的基礎が発生するのではないだろうか。

作業の手順

拙稿作成のために行った作業の手順を,再現性の意義も含めて示す。番号は本文中の手順番号と一致する。

利用した主なソフトウェアは次の通り。

R: R Core Team（2014）. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Cytoscape: Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T., Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research 2003 Nov; 13（11）:2498-504 URL http://www.cytoscape.org/

なお, > で始まる行は R での入力である。

手順番号: 1

連想辞典を用いて元のデータを作成する。

http://www.tesaurus.ru/dict/dict.php

に根となる語を入力して結果を得る。刺激語 быть を入れると

Статистика по запросу:

всего реакций на стимул: 509,

（中略）

Реакции	Частота
или не быть	45
человеком	44
жить	27
существовать	24
не быть	12
собой	12

（以下省略）の結果を得る。これをテキストエディタにペーストする。Реакции Частота から上の行は分析には不要。R と Cytoscape では,コメント行を # で指定できるので,不要な行の行頭には # を入れておくと良い。さらに,上記のウェブでここで反応語として刺激語を入力して同様の結果を得る。刺激語が返されなければ何もしない。完成したファイルを 根が быть なら“byt.txt”,рука なら“ruka.txt”,хороший なら“xoroshij.txt”として保存する。今回の分析では,根となる刺激語から連想語の結果を得て,その連想語をさらに刺激として連想語を得るという,いわば第 2 レベルまで作業を行った。

手順番号: 2

ruka モデルの作成には,“ruka.txt”の反応件数が多い上位 4 語,さらにこれら 3 語を刺激語として得られた語を使用した。ファイル名は“ruka_digest.txt”である。

R に入る。

> libary（igraph）

でこのパッケージを設定する。もともと入っていなければ別途導入しておく。テキストファイルをデータフレームとして読み込む。

> ruka.digest <- read.table（“ruka_digest.txt”, sep=“¥t”, comment.char=“#”）

このオプション部分は,ファイル名,テキストファイル区切り文字がタブであること,データとしては無視するコメント行の印として # を指定することを意味している。

次に,これをグラフの形式に変換する。

ruka.digest.g <- graph.data.frame（ruka.digest[1:2], directed=TRUE）

データフレームの 1 行目から 2 行目をグラフデータとして用いること,有向グラフであることを指定する。

そして,データが正しくグラフとして作成されたことを確認する。

> ruka_digest.g

IGRAPH DN-- 8 12 --

+ attr: name（v/c）

となる。D は有向グラフであること,次の数字はそれぞれ頂点の総数と辺の総数。

> E（ruka_digest.g）

Edge sequence:

[1] рука　-> нога

[2] рука　-> друга

（中略）

[11] левая　-> рука

[12] левая　-> сторона

で辺が求められる。

> byt <- read.table（“byt.txt”, sep=“¥t”, comment.char=“#”）

得られたデータフレームをグラフ形式に変換する。

> byt.g <- graph.data.frame（byt[1:2], directed=TRUE）

Cytoscape に入る。“ruka_digest.txt”をインポートする。この際,テキストファイルのオプションを表示して,コメント行として # を指定し,1 列目をソース,2 行目をターゲットにする。また Network Analyzer を実行する。ネットワークの図を表示する。画面下の表に,Node Table があるので,ここから必要な項目で並べ替える。可能なのは,InDegree（入次数）,OutDegree（出次数）,Betweenness Centrality（媒介中心性）など。

手順番号: 3

> betweenness（ruka_digest.g）

で求められる．媒介中心性は これを頂点の組み合わせ数 42 で割ると良い。

手順番号: 4

媒介中心性の大きな語は Cytoscape の Node Table を Betweenness Centrarlity を基準にソートすることで求められる。

手順番号: 5

任意の語が仲介となる連鎖を見つけることは難しい。次のように,語と語の最短経路を求めてみる。

> get.shortest.paths（byt.g, “я”, “счастливый”）

$vpath

$vpath[[1]]

[1] 72 22 64

すると,72, 22, 64 が頂点として得られる。これは頂点の番号なので,これを語にしてみる。

> V（byt.g）[c（72,22,64）]

Vertex sequence:

[1] “я” “человек” “счастливый”

である。また,任意の語を媒介する経路数は

> betweenness(byt.g, “человек”)

человек

37507.76

などで求められる。

手順番号: 6

> get.shortest.paths(ruka_digest.g, “рука”, “сторона”)

で,この 2 語の最短経路が求まる。返される値は

$vpath

$vpath[[1]]

[1] 1 3 4 8

のように頂点番号。この番号から実際の語を求めるには,

> V(ruka_digest.g)[c(1,3,4,8)]

Vertex sequence:

[1] “рука” “правая” “левая” “сторона”

とする。

手順番号: 7

平均最短距離を求めるには,手作業でなければ

> average.path.length(ruka_digest.g)

[1] 1.714286

とする。

手順番号: 8

> average.path.length(byt.g)

[1] 3.139026

手順番号: 9

次数分布のグラフを作成するには

> plot (degree.distribution(byt.g)[2:length (degree.distribution(byt.g))], main=“Degree Distribution of быть”, xlab=“Degree”, ylab= “Distribution”, xlim=c(1,500), ylim=c(0,1))

とする。main,xlab, ylab はそれぞれグラフのタイトル,横軸のラベル,縦軸のラベル。degree.distribution 関数では次数が 0 の度数は 2 番目のデータなので, 2 を指定する。また,degree.distribution（byt.g）は度数分布の長さ（大きさ）である。これを最後の値にするために指定すると良い。xlim と xlim はそれぞれ横軸の最小値と最大値。複数グラフの比較をするためにそろえることにした。

手順番号: 10
両対数グラフを作成するには
plot（degree.distribution（byt.g）[2:length（degree.distribution（byt.g））], main=“Degree Distribution of быть”, xlab=“Degree”, ylab=“Distribution”, log=“xy”, xlim=c（1,500））
とする。通常のグラフと異なる点は,log の指定である。ここでは,小数点以下の小さい値があるので,縦軸は 5e-02 などと表示される。これは, 5*10^-2,すなわち 5*10 のマイナス 2 乗 = 5*0.02 = 0.05 のことである。

参考文献

金明哲,『テキストデータの統計科学入門』,岩波書店,2009.

小林みどり,『あたらしいグラフ理論入門』,牧野書店,2013.

増田直紀・今野紀雄,『「複雑ネットワーク」とは何か―複雑な関係を読み解く新しいアプローチ』講談社,2006.

Aitchison J., Words in the mind : an introduction to the mental lexicon, 1994（『心のなかの言葉 -心内辞書への招待-』,宮谷真人・酒井弘訳, 培風館,2010）

Barabasi A.L., Linked: The New Science of Networks, Perseus Books Group, 2002（『新ネットワーク思考―世界のしくみを読み解く』,青木薫訳,NHK 出版 ,2002）

Caldarelli G. and Catanzaro M., Networks: A Very Short Introduction（Very Short Introductions）, Oxford University Press: Oxford, 2012（『ネットワーク科学―つながりが解き明かす世界のかたち』,増田直紀・高口太朗訳,丸善出版,2014）

Гак В.Г. Лексика как система // Энциклопедический словарь юного филолога, Сост. Панов М.В. Москва, 1984.

Караулов Ю.Н. и др. Русский ассоциативный словарь, Книга 1-2, Москва, 1994.

Рахилина Е.В., Когнитивный анализ предметных имен: семантика и сочетаемость, Москва, 2000.

PAC:「ロシア語連想辞典」Русский ассоциативный словарь, URL: www.tesaurus.ru/dict/dict.php.（2015年9月23日確認。この説明は URL: http://tesaurus.ru/ などにみられる。）

Уфимцева Н.В., Черкасова Г.А., Ассоциативная лексикография и исследования языкового сознания // Филология и культура（No.4, 38）, 2004.