

平成 29 年度 東京外国語大学オープンアカデミー
東京外国語大学語学研究所 企画
『コーパスから見えることば・文化・社会』
2017 年 10 月 24 日 (火) 第 3 回
「コーパスから見える 20 世紀韓国のことばと社会」
東京外国語大学准教授
南 潤珍

0. はじめに

皆さんこんばんは。東京外国語大学で朝鮮語学を教えている南（ナム）です。今日私がお話しするのは20世紀韓国のことば・社会についてです。始める前にお断りしておきたいのですが、もともとこの講座のテーマは、コーパスから見えることば・文化・社会だったのですが、講座をお引き受けしたときに具体的なタイトルをつけてくださいと言われてまして、自分は現代韓国の言語を研究しているので20世紀の朝鮮語に限って話すことにしました。私は、主に言語の分析をやってきましたので、文化や社会のことを深く取り上げるのは今の段階では無理があるのではないかと思います。主にことば研究の側面を話し、それが社会的にどのような意味を持つのかをこれからの研究の課題として考えていくという形でまとめたいと思います。

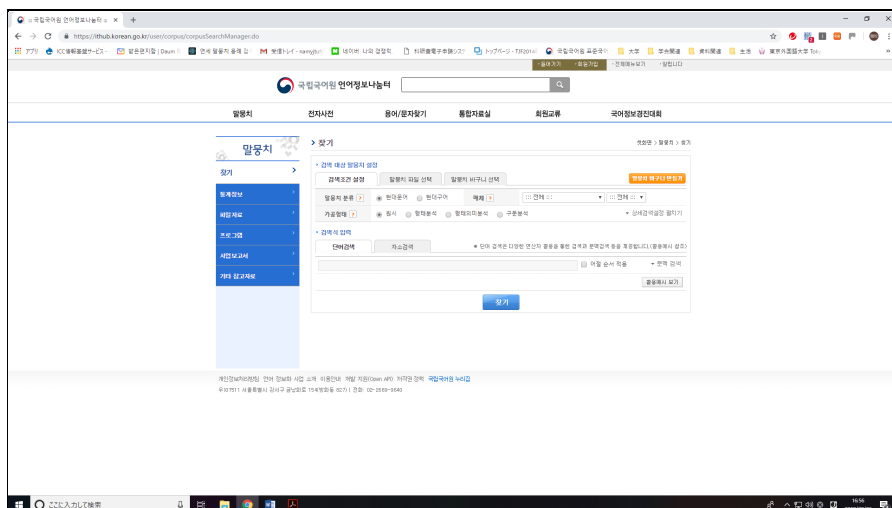
1. 韓国のコーパス事情

まず、朝鮮語のコーパスはどうなっているのかを見えます。コーパスについてはすでに2回の講座があったのでだいたい理解されていると思いますが、今日ご紹介する朝鮮語のコーパスは大きく3つに分けられます。「21世紀世宗計画」という韓国政府のプロジェクトで作られた「世宗コーパス」。延世大学言語情報研究院で作られた「延世20世紀韓国語コーパス」。そして高麗大学で作られた「SJ-RIKS拡張版」と「高麗話しことばコーパス」というものがあります。

1.1. 世宗コーパス(韓国文化体育観光部)

最初にあげた「世宗コーパス」は1998年から10年間、韓国政府文化体育観光部

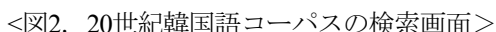
の支援を受けて作られたコーパスで、朝鮮語の様々なジャンルの書物や話しことばを集め、それをパソコンで処理できる形に加工し、検索ソフトなどと一緒に提供されています。ちなみに、世宗とは、韓国の文字ハングルを作った15世紀の王様の名前です。朝鮮語のコーパスとハングルという文字は近い関係にあるので、その名前をとって世宗コーパスとしたものでした。その構成は、書きことばコーパス、話しことばコーパス、そしてパラレル・コーパスと歴史コーパスとなっています。パラレルコーパスとは、同じ内容のテキストを異なることばで(朝鮮語と英語、朝鮮語と日本語のように)書いた/話した資料を集めたコーパスです。例えば、ハリー・ポッターだったら英語版と朝鮮語版、村上春樹だったら日本語版と朝鮮語版があるわけですが、そのようなテキストを内容に沿って、対応させながら並べておき、同じ内容のものを検索できる状態にしたものをいいます。歴史コーパスとは、現代語だけではなくハングルが作られた15世紀から19、20世紀初め頃までに作られた文献を集めて、そこに使われた朝鮮語を電子ファイル化したものです。そのような4つのグループに分けてコーパスを作りました。2002年、2007年、2009年、2010年に暫定版を公開し、最後に2011年に最終版を発表したのですが、その中で、書きことばコーパスが3,700万語くらい、話しことばコーパスが80万語くらい公開され発表されました。それらはDVDの形で配られましたが、現在は品切れとなり、Web上でダウンロードできるようになっています。公開されている資料のほか、もっとたくさんのコーパス資料が集められていると思われますが、著作権が未処理状態ものがあるということで、著作権の問題のないものだけでこの量になりました。



<図1. 世宗コーパスの検索画面>

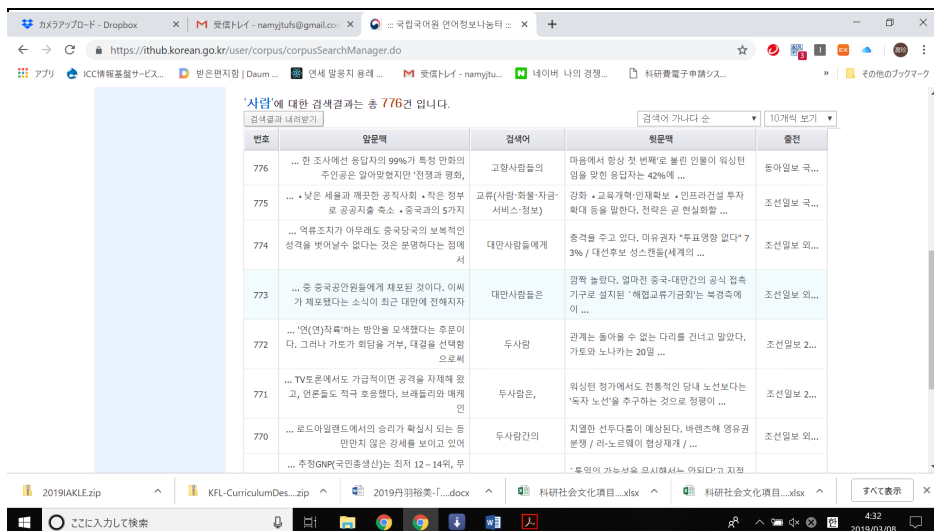
1.2. 20世紀韓国語コーパス(延世大学)

延世大学は、韓国でもっとも早い時期、すなわち1989年からコーパスを作り始めました。それは新しい朝鮮語の辞書を作るための基礎作業として始められました。従来の朝鮮語辞書は出版社で編集者が内省に基づいて作るものと考えられていましたが、この新しい辞書編纂の動きは、実際に使われた用例に基づき、言語学の研究成果を反映させた辞書が必要という認識から、延世大学の研究者や卒業生が立ち上げたプロジェクトです。この辞書編纂の過程で様々なコーパスの必要性に気づき、コーパスを作ってきました。朝鮮語学習者のための辞書である「연세한국어사전延世韓国語辞典」をはじめ、数種の辞書が刊行されておりますが、その過程で使われたコーパスの一つが、1900年から1999年まで出てきたテキスト（文献）を集めた「20世紀韓国語コーパス」です。新聞、文学作品、情報、単行本の文献、辞書、雑誌、教科書という部門に分けて全部で1億5000万語くらいのコーパスです。



世宗コーパスを高麗大学の研究員たちが修正または補充して作った拡張版で1億1600万語です。ですから公開されている世宗コーパスは話しことばで3700万だったんですが、実際は1億以上あって、それを高麗大学で手を加えて発表したものです。そして、今から話すのは公開されてはいないのですが高麗大学で作った話しことばコーパスです。これは1926年から1942年、そして1949年から1999年の2回、2つの時代に分けて昔のラジオ放送を録音したLPレコードや映画などの音声資料を集めてそれを文字化したもので、映画、ドラマ、漫画、普通学校読本を集めた資料です。ですが、これは未公開で、論文などで「こういう結果が出ました」という成果だけが発表されています。それが実際にどのような形になっているのか見てみましょう。

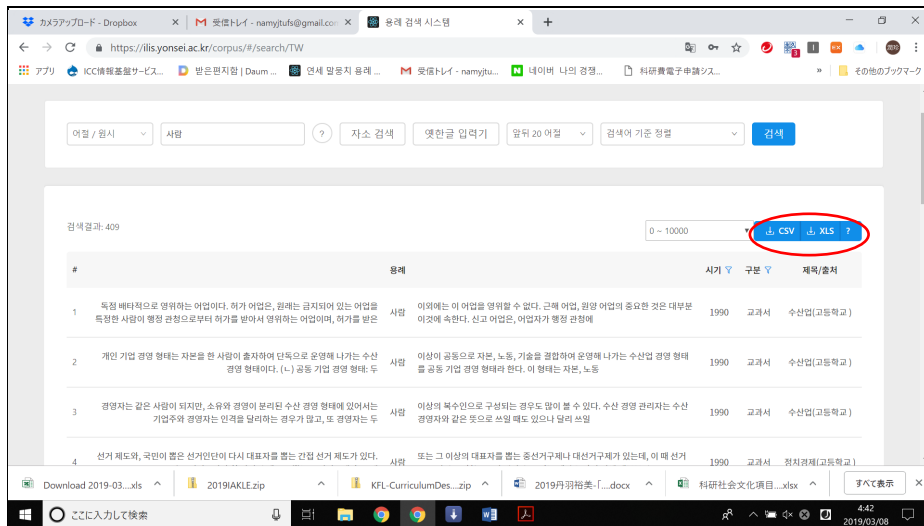
世宗コーパスはDVDという形で公開されていますが、今はWeb上でダウンロードもできますし直接検索もできるようになっています。時間がかかるので、あらかじめ私が検索しておきました。何を検索したかという、世宗コーパスの現代語の新聞の国際面に入っている「**사람salam**：人」という名詞を探してくださいと検索をかけて、その結果776の用例が出てきました。



<図3. 世宗コーパスの「사람salam:人」の検索画面>

この結果を見ると「사람salam: 人」の前の文脈と後ろの文脈で出典が出てきます。これはWeb上でも見られますがこれをダウンロードして自分のところでいろいろな処理ができるような仕組みになっています。

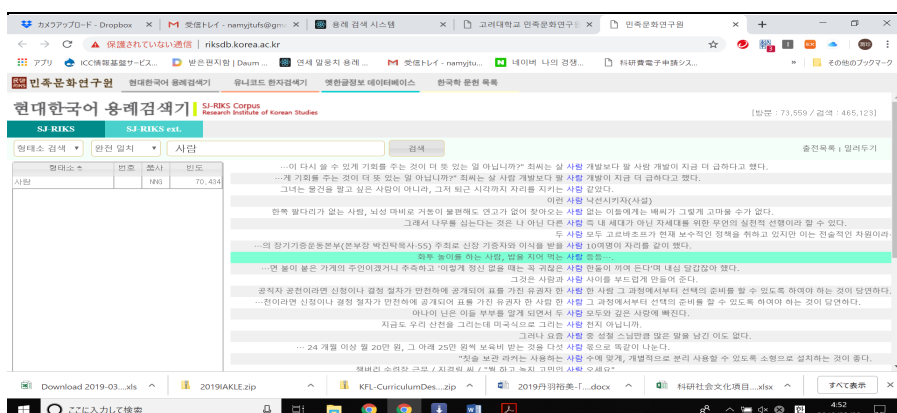
他のコーパスもだいたいこのような状況で、延世大学の20世紀韓国語のコーパスを見てみると、ここも同じく「사람salam: 人」という単語を検索したのですが、全部から検索すると1億5000万語から探すことになるので時間がかかります。今お見せしているのは、1990年代の教科書から出てくる「사람salam: 人」の用例409語の一部です。



<図4. 延世20世紀韓国語コーパスの「사람salam:人」の検索画面>

これもダウンロードできるようになっていますが少々ややこしく、10,000語単位でダウンロードできるようになっています。このように409語だといいいのですが、何十万、何百万の語が出てくる場合もあるので、それを10,000個単位でダウンロードするのは気が遠くなります。

そして、高麗大学の世宗コーパスを拡大したものから同じく「사람salam:人」を検索しました。

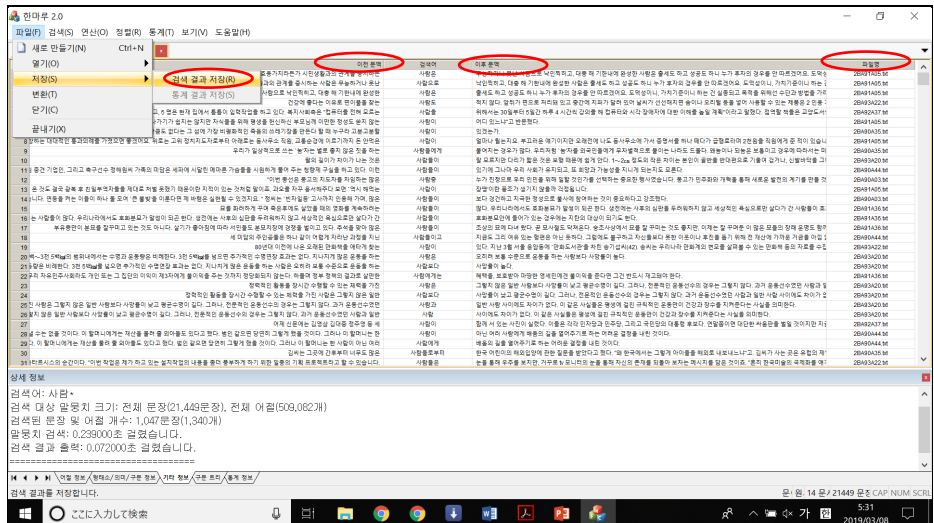


<図5. 高麗大学：世宗-民研コーパスの「사람salam:人」の検索画面>

このコーパスは、先程の世宗コーパスや延世大学の20世紀コーパスと違って、一度検索すると1億1600万を全部調べてその結果を出します。ですから、ここで「사람salam:人」は70,434個という結果が出ます。

問題なのはこの結果のダウンロードができないということです。検索結果から頻度を出したり、どういう状況でこのような単語が使われているのか、文脈を検討するわけですが、延世大学のコーパスも高麗大学のコーパスも検索結果をファイルの形で入手することが難しく、例えば延世コーパスだと10,000語単位で何度もダウンロードしているうちに、高麗大のコーパスの場合は、コピー・ペイストを何度も繰り返すうちに、そもそもなぜこんなことをやっているのだらうと考えることがよくあります。なぜコーパスを使って研究するのか？を考えると、自分が立てた仮説または「こういうものなのです」という主張の根拠としてコーパスでこういう結果を得ましたと出すわけです。もし自分が先程の高麗大学の1億以上を全部探して出した結果で「こういうことをやりました」と言っても、ある人は「小説は違うのではないか」「新聞では違う結果があるかもしれません」「もし1950年代だったら違うかもしれません」などという言われる可能性があります。ですから全部ひっくるめて探してもそれがどこからどれくらい例が集まったのかを提示しなければなりません。それなのに今のコーパスではそれがなかなか難しいです。そして頻度を出します。頻度は先程述べたようにパッと出てきますが、それが何を意味するのかを考えなければなりません。その頻度が全体でどのくらい占めているのかを言わないとその結果が使えないということです。例えば、1億1600の中で「사람salam:人」という名詞が70,000回出たとはいえませんが、名詞の中でどのくらいの割合で出てくるのかはわかりません。お見せしたWebで検索できるコーパスではそれが難しい状態になっています。それで、私の方でコーパスを使って研究するときは、Webで検索するのではなく、すでにあるコーパスのテキストから自分が必要なテキストだけを選んで検索しています。そのためにはいろいろな検索用のソフトウェアが必要なのですが、それがあのです。代表的なものが世宗コーパスの「한마루2hanmaru2」です。この検索道具を使うと、現代韓国語のコーパスに入っているテキストを選択して検索することででき、その結果をまとめてダウンロードしてエクセルで見ることができます。そして出てくる情報はここに出てくる前のキーワード、後ろのキーワードとファイルの名前があるので、このフ

ファイルの名前を知っていると、このファイルに何が入っているのかリストがあるのでリスト情報も提供されます。



<図6. 한마루2hanmaru2による世宗コーパスの「사람salam:人」の検索画面>

2. 頻度を利用した朝鮮語研究(1):コロケーション関係の調査

ここからはコーパスを用いて何を研究しているかについて、私自身の研究を中心に述べます。

まずは、日本語の「やさしい」という単語は「簡単な、親切的な」という意味の基本単語です。「簡単な」という意味ではなくて、人の態度などに対する「優しい」に対応する朝鮮語を辞書で調べてみると、次の6つぐらいの意味が提示されています。

(1) 「やさしい」に対応する朝鮮語の単語

- ① 부드럽다
- ② 착하다
- ③ 친절하다
- ④ 다정하다
- ⑤ 자상하다
- ⑥ 상냥하다

そうになると、朝鮮語を学んでいる日本語母語話者の学生は「やさしい」を朝鮮語でどう表現したらいいか迷います。この6つの使い分けができないと変な朝

鮮語になってしまいます。それを調べたい。でも辞書を引いてみると、①を説明するのに⑤を用いる、または、⑥の説明に③を用いるなど、延々とループする状態なので、辞書だけでは知りたい情報が得られないのが実情です。それを世宗コーパスから調べてみました。この使い分けを調べるために、①から⑥までの単語のコロケーションを調べました。コロケーションというのは、ある2つ以上の単語がよく一緒に現れて使われるもの、中には熟語や慣用句があるのですが、そうでないものも全部合わせて、いっしょによく使われるというのであればコロケーションと見做します。コロケーションの中でも①から⑥までの単語の後ろに何が出てくるかを調べました。

コロケーション	頻度	コロケーション	頻度
목소리 「声」	57	미소 「微笑」	17
것 「もの・こと」	35	곡선 「曲線」	16
맛 「味」	28	분위기 「雰囲気」	15
느낌 「感じ」	22	어조 「語調」	14

＜表1 ①「부드러운」のコロケーション：後項語彙（＋1位置）＞

表1を見ると①「부드러운」の後ろには「声」「感じ」「味」「微笑」「ことば」という単語が後ろに来ます。

コロケーション	頻度	コロケーション	頻度
사람 人	130	마음씨	14
일 こと	116	심성 心性	9
아이 子供	35	여자 女	8
마음心	19	것 もの	7
사마리아인 サマリア人	18	아기 赤ちゃん	6

＜表2 ②「착한」のコロケーション：後項語彙（＋1位置）＞

②착한の後ろにはいろいろなものが来ますが、「人」がよく来ます。「人」の中でも「子供」「女」「赤ちゃん」です。

コロケーション	頻度	コロケーション	頻度
설명說明	20	말씨 口調	4
태도態度	11	표정을	4
말 ことば	5	서비스 サービス	3
안내案内	5	안내서 案内書	3
마음気持ち	5		

<表3 ③「친절한」のコロケーション：後項語彙（＋1位置）>

③친절한の後ろには「説明」「態度」「案内」「ことば」が来ます。④의다정하다の後ろには「友達」「恋人」「隣人」「対話」など相手がいることを表す単語が後ろに来ます。⑤자상한の後ろには主に「人」が来ますが「人」の中でも「お父さん」「夫」「先生」「お母さん」「目上の人」などだいたい目上の人がよく使われます。⑥상냥한の後ろには「声」「微笑」もありますが「女性」「アナウンサー」「お嬢様」「案内状」という女性を表す単語が来ます。

これまでの結果をまとめると表4の通りです。

優しい	対象	関係性	人間（男女）	年齢
① 부드러운	無	無	—	—
② 착한	無	無	人間（女性）	下
③ 친절한	有	無	—	—
④ 다정한	無	有	—	—
⑤ 자상한	有	有	人間（男性）	上
⑥ 상냥한	無	無	人間（女性）	—

＜表4 「優しい」に対応する朝鮮語の単語の用法分類＞

①から⑥の中で、③と⑤はそういう態度を見せる対象が必要であり、④だけが相互の関係性を表します。即ち④は後ろに「友達」「恋人」「対話」のように2人の関係を前提とする単語が来ます。そして人間が後ろにくるのが②、⑤、⑥です。人間が来るとしても②の後ろに来る人間は女性、⑤は男性、⑥は女性が多いということがわかりました。そして、年齢や上下関係でみると②は目下、若い人、子供、⑤は目上の人である「お母さん」、「お父さん」、「先生」が来ます。例文をみてみます。「やさしい人だ。」の「やさしい」は、朝鮮語ですと男女で⑤か⑥かが分かれています。「やさしい彼氏だ」の場合は⑤を、「やさしい彼女だ」のときには⑥を用います。「兄のやさしい配慮」の場合、目上の人である兄に関するものなので⑤を使い、「いい子だった」は②を使います。②と⑤を見てピンと来ませんか？対象と関係性で違いますよね、②は対象や関係性がなく、⑤はあるという違いがあって、人間を表したものと結合する時、②は女性、⑤は男性とよく結合します。年齢については、②は目下の人、⑤は目上の人です。強いていえば、男性は上、女性は下、という考え方という分析ができます。「やさしい彼女」「やさしい彼氏」はどちらも「やさしい」ですが、彼女は⑥、彼氏は⑤。②と⑥は同じ女性を表しますが、②は子供と同じ扱いなので、ここから朝鮮語の底にあ

る男女に関する意識がちらっと見えてきます。そういう使い方をみると言えるでしょう。

3. 頻度を利用した朝鮮語研究(2): 否定文使用の推移

3.1. 朝鮮語の否定文と否定形式

頻度を利用した研究の2つ目です。20世紀の初めから終わりまでの変化を見るためのものです。日本語もそうですが、朝鮮語には「～しない」「～できない」という2つの否定があります。「～しない」を一般否定、「～できない」を能力否定といます。朝鮮語にも両方あるのですが、これから先が違います。日本語の否定は一般否定と能力否定において「～しない」「～できない」のようにそれぞれ一つの形で表します。動詞が前にきて形を変えて否定を作ります。それに比べて朝鮮語にはバリエーションがありまして、一般否定でも能力否定での2つの形があります。一般否定ですと、「～しない」が2つに分かれています。1つは動詞のあとに否定を表す否定述語(助動詞)が来るタイプです。もう1つは動詞の前に否定を表す副詞が来るタイプです。1つ目と2つ目の違いはあるのかについては、意見が分かれており決着はついていません。2つの否定文の意味が同じかどうかということは置いておいて、2つあるとしたら、なにか理由があるのではないかと考えるのが大きな前提で、どのように使い分けられるのかを知りたくなります。そういう問題が1つ。能力否定の場合も一般否定と同様に2つの形があります。述語や副詞は異なり、能力否定は能力否定の否定述語を用い、副詞についても一般否定文の否定副詞と異なる否定副詞を用います。結果的に4つの否定表現があるわけです。そこに第3の否定と言って「～するすべがない」に当たる表現が能力否定として使われます。つまり3つの能力否定があります。ですが3つ目は、昔の文献にはあまり出てこないもので、最近のものかもしれません。そのため、この第3の能力否定の表現はいつから、どのような場面で使われたのかを調べてみました。否定文の述語を使うものを①、副詞を使うものを②とし、どのような違いがあるか、そして能力否定の第3の否定と①番と②番の関係はどうか、頻度を用いて見てみます。

頻度を調査するコーパスとして、世宗コーパスの中から小説に限定して1900年代から1990年代まで10年区切りでできるだけ同じ量になるようにして表5のようにテキストを集めました。

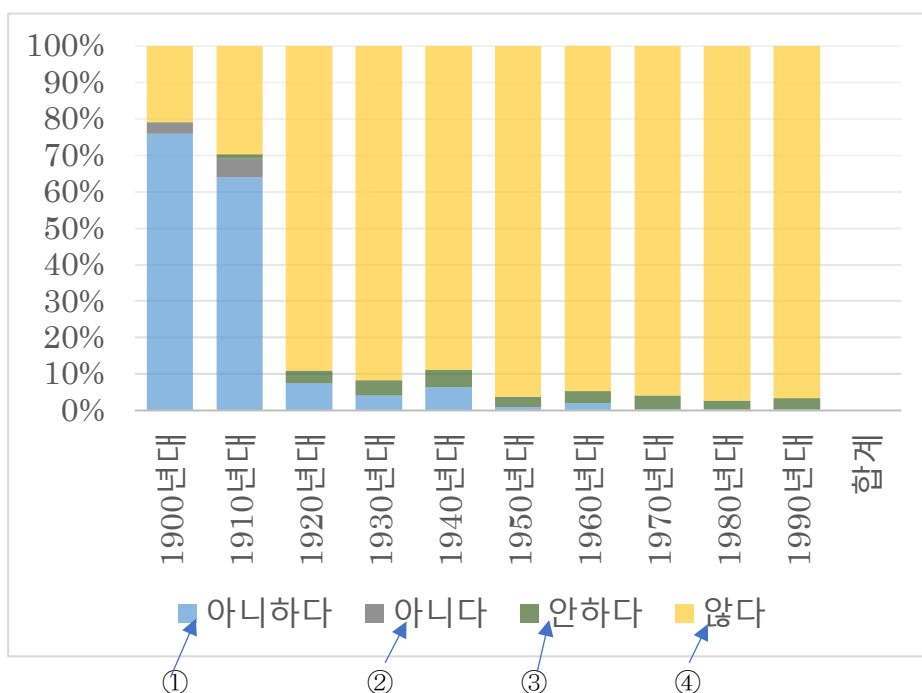
年代	作品種数	大きさ(語)	年代	作品種数	大きさ(語)
1900年代	13	200,070	1950年代	39	197,350
1910年代	14	193,534	1960年代	31	200,259
1920年代	51	200,108	1970年代	28	200,017
1930年代	42	252,555	1980年代	30	200,067
1940年代	30	147,905	1990年代	37	200,100
合計(全体)				315	1,991,965

＜表5. 20世紀小説コーパスの内訳＞

だいたい10年単位で20万語を目安に集めました。1990年代や1980年代になると入手できるものがたくさんありますが、1900年代や1910年代、1920年代になるとなかなか入手できませんでした。1940年代は意外と少ないです。ここから否定文を全部取り出しました。

3.2 否定述語の使用の推移

そして一般否定の否定述語の例をあげますと、20世紀初期に3音節の単語が少しずつ縮まって1音節の単語になってしまっています。今使っている1音節の否定述語は昔の文献には出てきません。15世紀にハングルが作られたのでハングルで書かれた文献をもっとも遡ると15世紀になるのですが、その時の文献から始めて見てもなかなか出てきません。17世紀18世紀あたりから少しずつ出てきて、3音節が2音節に2音節が1音節へと変化する過程を調べた研究によると、それが定着したのは19世紀頃であるそうです。でも、20世紀初め頃の1900年代や1910年代の小説を調べてみたら、①아니하-、②아니-, ③안하- ④않-の4つの形が出ました。①から④まで変化するなかで②、③が中間形として存在していたことがわかります。否定述語の形が1音節の④に定着したあとにも①や②、③も使われていたことがわかり、実状を調べてグラフにまとめてみました。



<グラフ1. 否定述語の使用頻度の推移>

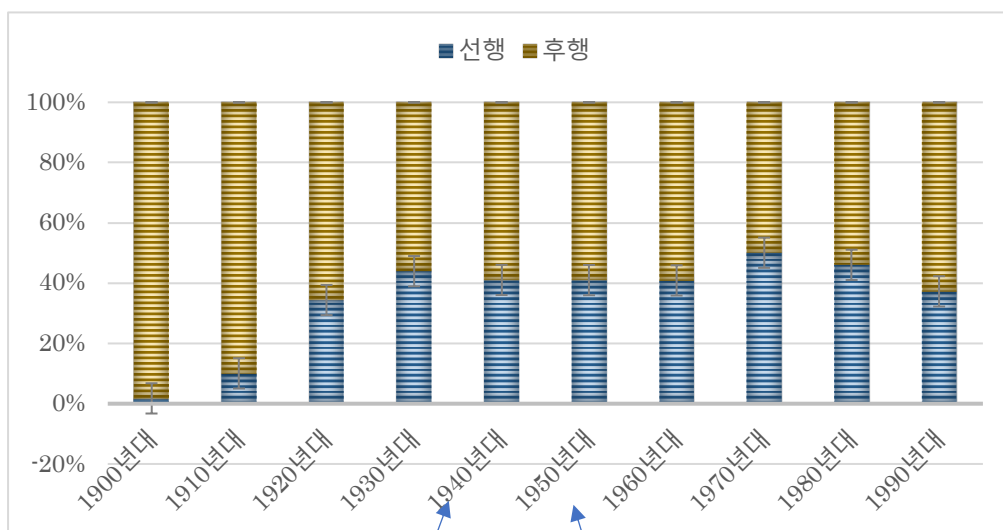
すると、①から④に変化するわけですが、1900年代をみると、①が948で、④は260で4分の1くらいになっています。すでにこの時期にも④が定着していると言えます。1920年代に①が急に少なくなっていて、1970年1980年1990年になるとさらに少なくなってきます。1900年代は④もありますが、一番多いのは①です。ここからわかるのは、①の頻度は1920年代に入って急激に減少、同時に④の頻度が上昇するという点です。これが何を意味するのかということですが、ここで韓国の小説事情が出てきます。古典小説は、昔話から形が洗練されてきて17、18世紀あたりからハングルで書かれ、読み物として読まれていた小説です。この古典小説は特有の文体があって、話しことばとはちょっと違うことばで書かれています。それに比べて、現代小説は、会話文と地の文が分かれており、地の文は書きことばで、会話文は話し言葉で書かれていることが多いです。古典小説から現代の小説に変遷していく過程でどんどん変化がありましたが、自然にそうなったのではなく、19世紀終わり頃に中国や日本を通じて西洋文学に接することができるようになった事実が背景にあります。特に日本の新しい小説に接した人が増えてきて、そういうものを作ろうとして小説を書いた人がいたのですが、いきなりそ

ういうふうにはなりませんでした。ですが、それは昔の古典小説とは違って、それを新小説と呼んでいました。新小説は新聞に連載され人気を博しました。なかには日本の小説を韓国風にアレンジして主人公も韓国人、場所も韓国になったものがあります。例えば『金色夜叉』です。新小説は1890年代から1910年代まで流行ってたくさん読まれました。実は1900年代1910年代にテキストとして入っているのは全部新小説です、それから、最初の近代小説というのは1920年代に書かれた短編小説から始まります。ですから、1920年代に急に④が多くなったのには文学及び小説出版の事情があるのです。19世紀に④として定着した否定述語は1920年代に入って安定した地位を確立しました、そしてそれには小説のジャンルのな変化も大きな理由がありました。

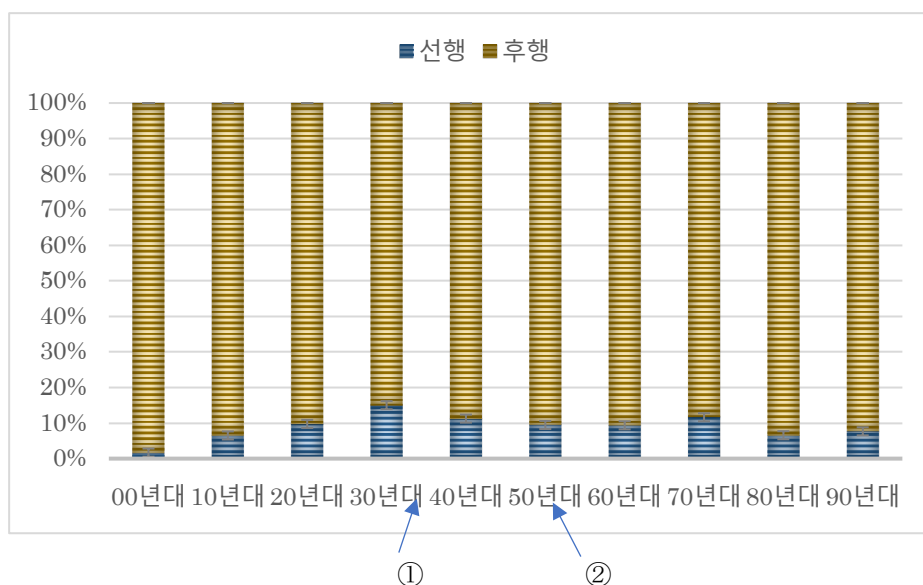
3.3. 一般否定文の2つのタイプの使用の推移

次に一般否定文がどのように使われたのかを見えます。地の文と会話文に分けて調べてみました。その結果をグラフに示したのがグラフ2とグラフ3です。①は否定副詞によるもの。②は否定述語によるものです。

地の文と会話文ともに②の否定述語による否定文の使用率が高いです。会話文では①と②の頻度の差が地の文のそれよりも縮んでいるのが特徴です。3.2で否定述語の形がどのように変化したのかを述べましたが、1920年代から会話文で①の副詞による否定文の使用が急激に伸びていることがわかります。1970年代1980年代にまた②の使用率が増えるのですが、その理由はわかりません。これからの課題です。強いていうと、②は話しことばでよく使われる形と言えるかもしれません。1920年代に②は会話文で急激に伸びていますが、これは会話文で話しことばで②を使う傾向が強まっており、それを作品に反映させるような小説の文体が確立されつつあるからと言えるでしょう。もしかしたら方言の影響があるかもしれないという話があります。方言によっては主に①を使う方言と②を主に使う方言があります。



<그래프2. 회話文における一般否定文の使用頻度の推移>



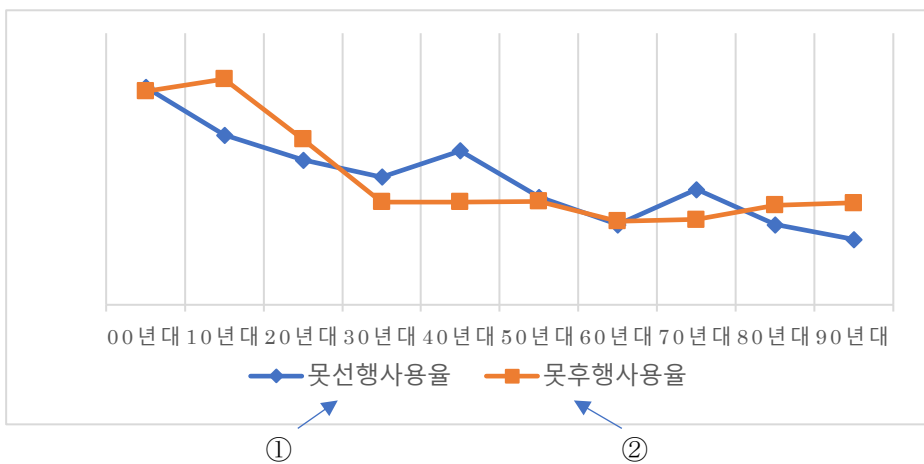
<그래프3. 地の文における一般否定文の使用頻度の推移>

1970年代以降、小説家の中では自分のことば即ち方言を解放させたと評価される人が現れますが、それは文学の方と連繫して調べてみないとわかりません。なんらかの方言との関係はあるかと思います。もう一つ調べなければならないのは、

新聞など話しことばが現れないもの、
放送やドラマがどうなっているのかもあわせてみる必要があるかもしれません。

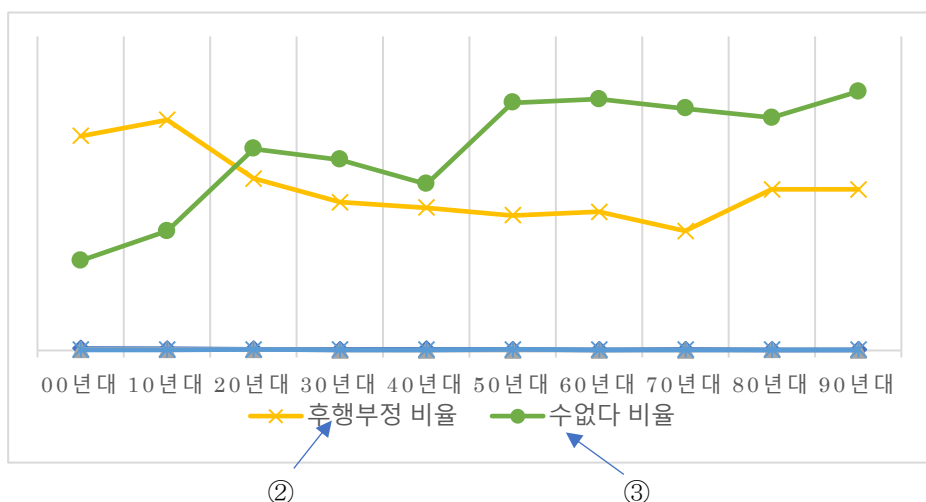
3.4. 能力否定文の使用の推移

「～できない」に関して①と②があり、さらに「～するすべがない」という③の形を調査した結果です。まずはグラフ4で①と②を比べて見ます。3.3でみた一般否定と異なり、はっきりとその傾向の違いは見えませんでした。



<グラフ4. 能力否定文の使用頻度の推移①と②>

一方、①と③、②と③をそれぞれ比べてみると、③を使う頻度が少しずつ増えていることがわかります。次のグラフ5は②と③を比べたものです。



＜グラフ5. 能力否定文の使用頻度の推移②と③＞

③は少しずつ増えて一般否定文の変化の時期と同様、1920年代に急に増えています。そして少しずつ減ってまた1950年代にぐんと伸びました。即ち、③は1920年代と1950年代にすごく伸びたことはわかります。なぜでしょうか。1920年代は一般否定文においても能力否定文においても変化を見せる時期です。1920年代は小説における文体お変化が起こり現代小説が定着した時期で、文学的要因で小説で朝鮮語の現代語の文体が定着したのと同じ流れで、そういう否定文の使い方が現れると言えます。

1950年代はなぜなのか考えたのですが、1945年にアメリカ軍が入ってきました。1948年までアメリカ軍がいて1948年に新政府が立ち上がり、1950年に朝鮮戦争が起こりまたアメリカ軍が入ってきました。韓国とアメリカは密接な関係になって今に至っていると思うのですが、1945年から1950年の間の出版事情を記述した本などによると、1945年から1950年は戦争が終わって物資が不足していて本が出せる状況ではなかったようです。その中でも、今とは規模は違いますが、教科書として翻訳書が多く出版されたそうです。このような事情から見ると、英語やドイツ語やフランス語といった西洋の本が翻訳されて出版されて、そのような翻訳文体に人々が親しんできたことが関係あるのではないかと考えられます。例えば、「できない」の前に形容詞はつけますか？形容詞の能力否定というのはそもそも矛盾です。形容詞は今の状態であって意志でそうなっているのではないです。青いのは青くしようとして青いのではなく自然に青いのであって、できる、できな

いの問題ではないです。ですが、朝鮮語でも同じで「～できない」という能力否定の①と②は形容詞には使えません。でも③は形容詞にと使えます。③はなぜ様々なところに、形容詞にまで付くことができるのかと考えたら、ある英語の表現が浮かび上がってきました。「It couldn't be better」(このうえなく良い)。英語の形容詞の比較級と不可能表現を合わせて最上級を表現しています。これを朝鮮語に訳すと③しかありません。ですから、これは仮説ですが、1950年代に③が増えたのはこのような文化的接触の影響があったからなのではないかとずっと考えてきました。

このように、頻度は、コーパスで現れた回数を数えて得られる単純な情報なんです。それを解釈する過程でいろいろな要因と結びつけるという作業をすること重要です。

4. まとめ

これまでの内容を要約します。まず、韓国のコーパス事情です。いろいろなコーパスが作られ公開されていますが、実際に研究に必要な頻度情報を細かくみるには Web 上に公開されていえるコーパスでは限界があって、世宗コーパスのようにもとのテキスト自体を利用することができるコーパスでないと検索以上のことには繋がりません。今の世宗コーパスでは DVD の形で公開することは終わっていて Web 上でダウンロードできるようになっています。それを利用して細かくみていくことができます。

そしてコーパスを用いて研究をすすめる中で常に考えなければならないのは、コーパスから何をみつけるのか、単に検索するだけでなく、それがどのような状況でどのようなテキストでどのように使われているのかを明らかにすることです。そしてその結果を通していろいろなことを説明するためには、その検索した結果を出したコーパスの構成はどのようなになっているのかが必要です。数字としてどれぐらいの母集団があってその中で検索したらどれぐらいの割合を占めるのかまで言えないと研究に繋がらないということです。頻度とことばの研究をするのですが、上述したとおり、使用できる・できないではなく、そういう傾向がある、たくさん使われている、あまり使われていない、比較的多い、比較的少ない、ということが言えるところに頻度に基づいた研究の意義が浮き彫りになります。それから、これがこうだから自分の頭の文がこういう構造だからこういう文

法検証はこのように説明できる、私はそう思わない、それは成り立たない、それですぐ論争になるのが昔の研究でした。そうではなくて実際どのように使われているのかを客観的な証拠をもってことばの研究ができるようになります。頻度から見えてくることはいろいろあるのですが、今までは頻度自体を求めることが大変な作業でしたのでそれに力を入れていたのですが、ある程度コーパスも出ていますし検索ソフトもあるので、いろいろと便利になってきています。頻度を得るのが以前と比べてできるようになったので、そこからもう一步踏み込んで、その頻度が意味することは何か。例えば「やさしい」に対応する 6 つの単語の中で、頻度的に見てこういう傾向があってこれはどういう意味なのか。もちろん朝鮮語を勉強する人がその使い方を覚えて正確に使えるということもあるのですが、もう一つは、その言語現象に現れる言語を使う人の考え方というのものもある程度見えてきます。

そして最後に、ことばだけではことばが見えないということです。なぜ 1950 年代に頻度が上がったのかはいくらグラフや表を見てもわかりません。もっと細かく見る必要があることと、細かく見るとして何を見るかというのを決めるためには、そのことばが使われた時期の社会的文化的状況をふまえる必要があります。そして視野を広げてことばだけでなく社会や文化と繋がればもっと見えることがあるということを今日はお伝えしたいと思います。(完)

コラム 十韓国の人名

日本や中国と同じく、韓国の人名は、苗字と名前で構成されます。

苗字は(韓国では「성 seong: 姓」という。以下「苗字」)漢字 1 字のものが多く、南宮、獨孤のような 2 文字のは珍しい方です。

韓国では、苗字は変えられないものとなっており、夫婦の苗字が異なることはごく自然で一般的ことです。昔は、子供は父の苗字を継ぐものと思われていましたが、2008 年からは、夫婦の合意の下で母の苗字を継ぐこともできるようになりました。

韓国の苗字は元々親族関係を表すものでしたが、地名に由来する「본 bon:本」によってより細分化され、同じ苗字の同じ本は「宗会」という組織の元で一つの家を成します。例えば筆者の苗字「南」には、朝鮮半島南部の地名でもある「英陽、宜寧、高城」の 3 つの「本」があり、「英陽・南」と「宜寧・南」は苗字は同じであっても別の家となります。「宗会」は、薄れてはいますが、まだまだ大きな影響力を持っています。

その一つが、名前です。韓国人の大部分は「苗字 1 文字に名前 2 文字(漢字)」を持っていますが、その 2 文字のうち 1 文字は、「宗会」によって数十代分まで定められている漢字を用います。この 1 文字は、一般的に、陰陽五行の五行(木・火・土・金・水)の字を含む漢字となっております。例として、ある親子 6 代の名前を挙げてみますが、「相佑」さんの子供は「光炫」さん、その子供は「基浩」さんそして「真銓」-「潤美」-「景植」のように代別に五行の字が入っています。おそらく「景植」さんの次の代の人の名前には、「火」の入った字が名前の前部に来るでしょう。初めてあった人でも、同じ本の同じ苗字であることが分かれば、名前を手掛かりにお互いの関係性が大体分かるようになります。

ところが、最近は、漢字を用いず、ハングルでしか標記できない名前が増えていきます。こうした名前は 3 文字以上のものも多く、これからはより多様化していくことが予想されます。

朝鮮語を知るための3冊

・ * ・ ... † ... ・ * ・ * ・ ... † ... ・ * ・ ... † ... ・ * ・



李翊燮他（著）／梅田博之（監修）前田真彦（訳）
「韓国語概説」大修館書店 2004 年

韓国の国語学者 3 人による、非朝鮮語母語話者のための解説書です。朝鮮語の特徴を文字・音韻から単語・文構造や敬語、語史および方言に分けて分かりやすく説明しています。

... * ————— *

趙義成訳注「訓民正音」東洋文庫 800 平凡社 2010 年

朝鮮語の文字ハングル(訓民正音)についてその創制の当事者によって書かれた解説書の日本語訳に原本の影印及び詳細で的確な注釈を施した本です。ハングルのしくみだけでなく、その哲学的・言語学的基盤を知ることができる本です。



... * ————— *



韓美卿・梅田博之「韓国語の敬語入門」大修館書店 2009 年

韓国出身の日本語学者と日本の朝鮮語学者による解説書です。韓国ドラマで使用されている例文や題材を扱っているため、朝鮮語学習書としても活用できます。

... * ————— *