

国立国語研究所学術情報リポジトリ

世界の言語研究所（17） 延世大学校言語情報研究院（ILIS）と韓国語辞典の編纂（韓国）

メタデータ	言語: Japanese 出版者: 国立国語研究所 公開日: 2019-03-25 キーワード: 作成者: 徐, 尚揆 メールアドレス: 所属:
URL	https://repository.ninjal.ac.jp/records/2157

延世大学校言語情報研究院 (ILIS) と韓国語辞典の編纂 (韓国)

ソ サン キュ
徐 尚 揆 (韓国・延世大学校)

1. はじめに

延世大学校言語情報研究院 (ILIS: Institute of Language and Information Studies) を特徴づけるキーワードとして、何よりもまず思い浮かぶのは、「辞典」、「コーパス」(韓国語では「マルムンチ 말문치」「マルモドゥム 말모듬」(「ことばの集合体」の意))、「コーパス言語学」、「国語情報学」(Korean Informatics)といったことばである。本論では、これらのキーワードを中心に、ILIS の主な活動と、1980年代後半以降 ILIS で活発に行われた新しい方法論による辞典編纂について紹介する。この新しい方法は、辞典の編纂に大規模な言語資料と各種の電算技法を活用するもので、それまでの辞典編纂方法とは根本的に異なるものである。

ILIS の設立は非常に特殊な経緯をたどった。延世大学校が設立100周年を迎えた1986年の初めに、延世大学校文科大学のそれぞれ専攻が異なる数人の教授が、大学200周年に向けた文化事業として『韓国語大辞典』の編纂を提案した。彼らは自発的に辞書に関する共同研究を始めるとともに、大学の内外に向けて辞典編纂の必要性を訴えた。その結果、その年の10月に延世大の教授290人の連名で「韓国語辞典編纂会」が組織され、翌1987年には国内初の辞書学専門誌である『辞典編纂学研究』が同編纂会によって創刊された(この雑誌は2003年に『言語情報と辞典編纂』と改称され、現在13集まで刊行されている)。編纂会の指導者たちは、月例の共同研究会を開催する一方で、私費でヨーロッパやアメリカなどの辞典関連の学会に参加したり、著名な辞書出版社を訪問したりした。1980年代後半のこのような努力が、ILIS 設立の直接の基盤となった。

この活動は、当時はソウルの一私立大学のささやかな試みと映るだけだったが、その後10年も経たないうちに、韓国語辞典の編纂史に残る様々な変化をもたらした。新しいコンセプトにもとづく辞書の出版、韓国におけるコーパス言語学(マルムンチ言語学)、国語情報学という新しい研究領域の開拓など、当時は想像もできなかった変化がここから始まったのである。

大学当局は、これらの努力を支援するために、1986年に大学附設の研究所である「韓国語辞典編纂室」の設置を認めた。それ以後、2度の名称変更とともに規模も拡張され、現在では「言語情報研究院」(ILIS)の名称で、延世大の三大“校策研究所”(大学本部が直接運営を支援する研究所)の一つとなっている。2005年現在、ILIS は、辞典研究センター、言語情報研究センター、地名及び歴史資料研究センター、音声分析研究センター、専門用語研究センター、外国語としての韓国語研究センター、情報技術研究センターの七つのセンター、ならびに大学院言語情報学課程を運営している。

2. 主な韓国語コーパス

韓国語のコーパスが本格的に構築、活用されたのは、ILISの前身である「韓国語辞典編纂室」の「延世コーパス1」(Yonsei Corpus 1)が最初である。以後、1990年代初頭には、国立国語研究院(現在の国立国語院)、高麗大学校民族文化研究院、韓国電算院(KAIST)などが、競って大規模コーパスの構築に着手した。そのうち、国立国語研究院は『標準国語大辞典』(1999年刊行)を、また高麗大学校民族文化研究院は『韓国語中辞典』(未刊)の編纂を目標にしていたという点で、延世コーパスと似たところがある。

これらのコーパスは、1998年より政府の支援で行われている「21世紀セジョン(世宗)計画」を中心として統合されるとともに、新しい資料が引き続き開発されている。初期には主に書き言葉中心のコーパスが構築されたが、1990年代後半から話し言葉のコーパスがILISの主導により本格的に構築され始めた。

以下、大学研究所や研究機関によって作られた主な大規模コーパスを紹介する。

「延世(韓国語)コーパス」

ILISにおいて1987年から現在まで継続的に構築されているコーパスで、1987～1996年にかけて収集された「延世コーパス1～9」(約4300万語節:「語節」は日本語の「文節」にはほぼ相当)と各種の特殊コーパスで構成されている(詳細は後述)。「延世コーパス1～9」は書き言葉を中心にしたものだが、話し言葉も一部含まれている。特殊コーパスには、形態素注釈コーパス、意味注釈コーパス、多国語並列コーパス、韓国語学習者誤用コーパスなどがある。これらのコーパスは、辞典編纂ならびに国語学・国語情報学の研究に活用されている。

「延世コーパス1～2」(それぞれ300万語節、110万語節)は、韓国語初の均衡コーパス(balanced corpus)である。当時は、corpusを翻訳して「マルムンチ」という新しい用語を作り出すことから始めて、コーパスの収集方法、収集テキストの選定方法の決定、コーパスの規模設定など、すべて一からやらねばならなかった。そのことを考えると、この二つの均衡コーパスの構成に関する議論と実際の構築は非常に意義深い試みであった。

今日では、「21世紀セジョン計画」において1000万語節規模の均衡コーパスが構成されたことや、コンピュータ環境の変化などにより、コーパスの均衡性や代表性に関する議論はますます少なくなっている。しかし、個人レベルの研究環境にあっては、多様な目的と方法論による均衡コーパスの構成は今でも重要な問題と言える。

「国立国語研究院コーパス」

1992～1999年にかけて、国立国語研究院によって構築された合計6765万語節のコーパスである。1400～1990年代の書き言葉を中心に構成されており、カバーされている時代は最も広い。話し言葉もごく一部含まれている。このコーパスは『標準国語大辞典』(国立国語研究院、1999年)の編纂を主な目的として構築されたものだが、各種の国語研究に活用されている。

「韓国語コーパス1」(韓国語では「韓国語マルモドゥム1」)

辞典編纂と国語研究を目的として、高麗大学校民族文化研究院により1995年に構築されたコーパス(約1000万語節)である。主に1910～1990年代の書き言葉の資料で構成されているが、話し言葉もごく一部含まれている。

「国語情報処理基盤構築コーパス」

国語情報処理の技術開発のために、1994～1997年にかけて、政府の支援のもとで韓国科学技術院(KAIST)が構築したコーパス。書き言葉の資料約7158万語節からなるが、1998年に「21世紀セジョン計画コーパス」に吸収された。

「21世紀セジョン計画コーパス」

「21世紀セジョン計画」は、政府の支援のもとで1998年より実施されている計画である(2007年に終了予定)。その基礎資料構築部門では様々なコーパスが開発されている。コーパスの構築は、国家レベルの統合コーパスの構築を目標にして、ILIS(主に話し言葉と多国語並列資料)と高麗大学校民族文化研究院(主に書き言葉資料)を中心とする共同研究グループにより実施されている。このコーパスは、今まで韓国で構築されたコーパスの中で最も大規模かつ広範囲なものである。

2004年末現在、書き言葉資料は、原始コーパス(タグなしのテキストコーパス)6000万語節、形態素分析コーパス1220万語節、語彙意味注釈(同音異義語区分)つきコーパス312万語節、構文分析コーパス36万語節、韓英・英韓並列コーパス400万語節などとなっている。話し言葉資料は、原始コーパス333万語節、形態素分析コーパス68万語節となっている。その他、北朝鮮及び海外韓国語コーパス、歴史資料コーパス、口碑文学コーパスなどが構築されている。

3. 「延世コーパス」の概要

「延世コーパス」(YSC:Yonsei Corpora)は、『延世韓国語辞典』の編纂のために構築されたもので、1987年の「延世コーパス1」(約300万語節)を皮切りに、1996年までに構築された9種のコーパスの総称である。

○均衡コーパス(balanced corpus)

「延世コーパス1」(300万語節)

「延世コーパス」の最初のコーパスである。サンプルの代表性の確保、ならびにサンプルの選定基準をつくるために、現代韓国人(専門家、一般市民の2グループ)の読書に関する実態調査にもとづき構築された。このコーパスは、各類目別に一定の比率によって選定されたテキストからサンプルテキストを一定量ずつ抜き出した均衡コーパスである。

「延世コーパス2」(110万語節)

このコーパスでは、図書貸出頻度にもとづく均衡コーパスを構築するために、主題別選定基準を採用した。文献情報学的方法を応用し、可能なすべての主題と分野の単語が網羅されるように、デューイ十進分類法(Dewey Decimal Classification)にもとづき、韓国語文献を大きく10のカテゴリーに分けて収集した。また、図書貸出頻度は各単語の認知度を間接的に反映するという仮定のもとで、主に1987～1988年の図書を対象として、貸出頻度が高い資料を中心に234個のサンプルを選定した。主題別には、総類(7.8%)、哲学(9.9%)、宗教(10.7%)、社会科学(12.8%)、言語(5.7%)、純粋科学(11%)、応用科学(11.7%)、芸術(8.1%)、文学(11.2%)、歴史(11.3%)などで構成されている。

○時代別コーパス

「延世コーパス3：1980年代文献」(598万語節)

このコーパスでは、人々がより多く接する文献が各単語の認知度を間接的に反映するという仮定のもと、1980年代優秀出版物目録にもとづいて、サンプルが選定されている。

「延世コーパス4：1990年代口語」(77万語節)

1995年に構築されたコーパスで、実際に使用された話し言葉を録音して転記する「純粋話し言葉」と、戯曲、放送台本、シナリオなどの「準話し言葉」により構成される。具体的には、対話(26%)、講演(24%)、相談(14%)、戯曲・台本(13%)、DJ放送(13%)、討論(8%)、会議(2%)などからなる。ここには、発話者の年齢、性別、職業に関する情報、発話者の数、発話の性格、転記者情報、録音時間情報などが記載されている。

「延世コーパス5：1970年代文献」(860万語節)

1970年代の文献を対象にして、教科書から新聞まで多様な文献資料から構成されている。具体的には、新聞(10%)、小説・随筆(50%)、一般書籍(35%)、教科書(5%)などの文献資料が収集されている。

「延世コーパス6：1960年代文献」(723万語節)

解放以後の国語の姿を直接反映したコーパス構築のため、1980年代以前の時代を10年単位でさかのぼる形でコーパスの構築を開始した。当初は1000万語節を目標にしていたが、現在は中断状態にある。その理由としては、この時期の文献にはハングルと漢字が混じった文章が多く、実際の入力過程で様々な困難が生じたことなどがあげられる。

「延世コーパス7：1990年代文献」(1367万語節)

1990年代中盤までの文献資料を代表する資料として収集された。1994年から1995年にかけて構築されたので、実際には1990年代中ごろまでの文献を代表するものと言える。

「延世コーパス 8：教科書コーパス」(87万語節)

初等学校全科目と、中・高等学校の国語・社会を中心に構成され、第5次、第6次教科課程の二つの種類が収集された。

「延世コーパス 9：児童図書コーパス」(150万語節)

純粋な国語の使用頻度が高く、文章も国語の構造に比較的合致する、児童教育用図書としてふさわしい資料を選定して、1996年に構築された。

「延世コーパス5, 6, 7」は、ほぼ同一の構成方法によって構築され、総類、哲学、宗教、社会科学、芸術、文学、歴史学などのサンプルで構成されているが、全体としては、文学と社会科学の分野が圧倒的多数を占める。

このほかに、1998年の『延世韓国語辞典』の出版以後に各種研究などの目的のために新たに構築された特殊コーパスには、次のようなものがある。

「現代韓国語話し言葉転記コーパス」(原始333万語節、形態素分析68万語節)

自然発話を中心とした話し言葉を録音、転記したコーパスであり、「21世紀セッション計画」コーパス構築の一環として1998年から構築された。放送資料まで含めると、話者の数は延べ6337名にのぼる。独話(27.7%)と二人以上の対話(72.3%)から成る。独話資料は、公的独話(約23%：講義、講演、発表、自己紹介、業務報告、業務プレゼンテーション、演説、説教など)と私的独話(約4.6%：経験談、映画や小説などの筋書、電話メッセージなど)から構成されている。対話資料は、公的談話(約30.5%：討論、相談、会議、求職インタビュー、面接、教室対話、購買対話、診療対話、公共機関での対話など)と私的対話(約41.8%：日常対話、食事対話、電話対話など)から構成されている。これ以外に、2003年には大学生話し言葉転記コーパス(約15万語節)が構成され、音韻、形態、語彙、文法、語用などに関する総合的考察が行われた。

「引用句コーパス」(350万語節)

『現代韓国語文章引用句辞典』(仮称)を編纂する目的で、1910年代から1970年代までの主要文学作品を選定して構築した。

「現代詩コーパス」(85万語節)

「韓国代表詩人100人選集」を入力した、韻文だけで構成されたコーパスである。

「基礎学習コーパス 1」(177万語節)

『初等国語辞典』(2002年)の編纂を目的に構築されたコーパスである。教科書(5%)、児童新聞(10%)、児童雑誌(10%)、百科事典(15%)、児童創作作品(5%)、マンガ(5%)、童話・小説・伝記(35%)などの書き言葉と、児童放送用シナリオ(10%)、その他(5%)など

から構成されている。

「韓国語教育コーパス1」(100万語節)

外国語としての韓国語教育のための基礎語彙の選定、意味注釈、さらには『外国人のための韓国語学習辞典』(2003年)の編集を目的に収集された。主に韓国語教育に必要と判断された小説、随筆、手記、紀行、論説などの一般教養テキストと、韓国語教材、初等教科書(国語、社会、生活科)を中心にして、話し言葉と準話し言葉を含めている。以後、この資料を基盤として、形態素分析結果を付与した「韓国語教育コーパス2」、『延世韓国語辞典』の意味記述を基準にして基礎語彙5000個に意味標識を付加した「韓国語教育コーパス3」などが作られた。

「韓国語学習者コーパス」(約66万語節)

2001~2002年にかけて、韓国の主要教育機関で韓国語を学ぶ学習者の対話と作文を収集したものの。話し言葉資料は、主に討論と対話(教師対学生、学生対学生)、絵を見ての説明、本や映画などの筋書の発話などからなる。また、作文資料は、自由作文、課題作文、試験作文、作文コンテストの作文、論文、報告書などで構成されている。12個の収集群に計3952個のサンプルが収集された。

4. コーパスを利用した辞典編纂

前述のように、ILIS設立の目標は『延世韓国語大辞典』の編纂である。その目標と精神は、『延世韓国語辞典』の序文に端的に表されている(同辞典のインターネット検索サービスのホームページ <http://dic.dreamwiz.com> に掲載されている「延世韓国語辞典の編纂の歴史」より引用)。

この世で私たちの言葉が使われ始めてから数千年が経ち、1943年朝鮮語学会(現ハングル学会)により『朝鮮語大辞典』(現在の『国語大辞典』)が完成したおかげで、私たちの言葉と思想の材料となる要素のリストが初めて詳らかにされたが、外勢強占[注:外国による侵略]と同族相残[注:同じ民族どうしの戦争]により、同辞典は1957年に刊行を終えた。それ以後の辞書は、ほとんどがこの辞典に加筆や部分的な修正を行ったものである。

1960年以来私たちの力量の急激な伸張とともに、私たちの言葉も大きく発展した。そういう変化を既存の国語辞典が単純な加筆で表すことはできなかった。まったく新しい国語辞典の編纂が必要であった。(中略)

私たちは、研究を通じて、今日の辞典編纂には大規模な「コーパス」の構築と「コーパス」の電算処理が必要なことを痛感した。「コーパス」とは、一つの言語で書かれた文と言葉をできるだけ多様な形で集積し、大容量の電算機で処理したものである。数年間の苦勞の末、私たちはついに韓国初の大規模現代韓国語コーパスを構築し、それを処理するための各種ツールを開発した。この過程で、私たちは1956年以後初めて現代国語の語彙頻度を調査

し、国内に「コーパス」と「コーパス言語学」の概念を導入して、国語学の新しい世界を切り開いた。(以下略)

今日の時点で見ると、ILIS の最も大きな成果は、言うまでもなく『延世韓国語辞典』の編集である。2144ページの分量に約52000前後の見出し語が含まれるこの辞典は、語数の面で見れば、中辞典と同じ程度である。また、商品としての価値という観点からは、当初はあまり大きな成功を期待することはできなかった。学術的水準の高さがそのまま商業的成功に結びつくものではなく、韓国の辞典市場も日々縮小していくという状態だったからである。その上、辞典需要の大多数が学生であり、辞典市場の主力は英語を中心にした学習辞典であった。

それにもかかわらず、この辞典が注目されたことにはいくつかの理由がある。

第一に、この辞典が、『韓国語大辞典』の編纂という遠大な作業の中間段階で発刊されたものであり、大辞典編纂のための研究と努力が1980年代中盤から現在に至るまで、その目標を失わずに持続されていることである。

第二に、この辞典の開発の過程で、その以前にはなかった新しいコンセプトと方法論が応用されたことである。この辞典によって、韓国語辞典の編纂史上初めて、大量の言語資料の電算化と語彙分析を基礎とした辞典編集の方法が確立された。語彙記述の方法も従来の辞典とはまったく異なるものであった。

第三に、この辞典の編纂過程を通じて、韓国にコーパス言語学的な研究方法が導入され、定着し始めた点、そして、言語学、情報学、電算学など多様な学問分野の統合を目指す「国語情報学」という新しい学問領域が成立した点である。

『延世韓国語辞典』の、従来の辞典と異なる特徴としては、以下の点があげられる。

- (1) 見出し語の選定を含めて、意味の定義、用例の選定、文法情報などのすべての記述が、コーパスに出現した分布と実際の用例分析を通じてなされている。見出し語選定のために3回にわたって語彙頻度を調査し、最終的に「延世コーパス4」を除く延世コーパス全体を調査し、そのうち、頻度14回以上の語彙(5万余語、全体での頻度占有率約87%)を記述対象とした。
- (2) コーパスを活用するための辞典編纂ツール(用例検索、原稿の電算的作成、原稿データベース管理、辞書原稿変換など)を積極的に開発し、イントラネット環境での統合的な辞典編纂環境の構築を試みている。結果的には、これらのツールは個別のプログラムとして活用されるにとどまり、一つの辞典編纂システムとして統合することはできなかった。特に辞典編集プログラムは、辞典の構造の複雑さと時間不足により、結局は完成することはできなかった。今日の観点で見れば中途半端な成功と評価される程度の水準であろうが、辞典編纂の方法論の開拓という観点から見れば、非常に意義深い試みであったと言うべきである。公式にはILISの三つ目の辞典になる『延世現代韓国語大辞典』では、1945年から現在までの韓国語を「現代国語」として定義し、この時期の資料に現れる約15万個の単語収録を目標にしている。以前とは違い、この辞典の場合は、初期段階ですでに用例検索プ

ログラムと辞典編集プログラムが開発されているが、それも『延世韓国語辞典』におけるノウハウの蓄積があって初めて成り立っているものである。

- (3) 述語の見出し語を提示する形式を既存の辞典とは異なるものにした。例えば、用言を見出し語として提示する際には、語尾「다(da)」の部分小さな字で印刷した。これは、活用語尾が出現する位置を明確に示し、その位置に多くの語尾がつくという特性を表すためである。
- (4) 文法及び用法の説明のために、参考欄に格フレーム(case frame)と結合価を示し、用言が取る文型を一目で理解できるようにした。これは既存の辞典では重視されてこなかったことである。そのほか、参考欄には外来語、連語、類義語、反意語、関連語などに関する詳細な情報を示した。
- (5) 見出し語の意味を説明する際には、必ず実際の用例を提示した。用例は原則としてコーパスから抜き出したものである。しかし、実際には、コーパスに現れた文章をそのまま引用できない場合が多く、多くの場合は unnecessary 部分を削除するなどの修正が必要であった。以上に述べた辞典編纂の目的と方法論は、その後刊行された成果物においても一貫している。その主なものは以下のとおりである。

- ・『延世韓国語辞典』((株)斗山東亜, 1998年)
(インターネット検索サービス(2000年): <http://dic.dreamwiz.com/krdic/>, ブリタニカ百科事典韓国語版に収録(2000年): <http://www.britannica.co.kr/>)
- ・『東亜延世初等国語辞典』((株)斗山東亜, 2002年)
- ・『外国人のための韓国語学習辞典』(ソ サンギョほか編, 文化観光部・韓国語世界化財団, 2003年)
- ・『韓国語学習基本語彙意味頻度辞典』(ソ サンギョ編, 2005年予定)

最後にあげた『韓国語学習基本語彙意味頻度辞典』は、約100万語節のコーパスに現れた各基本語彙の用法を『延世韓国語辞典』の意味記述によって分類し、注釈を加えたもので、同音異義語の区別を越えた多義的意味の項目区分が試みられている。2005年3月現在、約5200余りの語が説明対象になっている。

5. 言語研究と国語情報学の先端的研究領域の開拓

『延世韓国語辞典』の編纂は延世大学校の国語学研究に大きな影響を与えた。実際の言語資料を言語学的に分析し、辞典で記述するという方法論を経験した人々を中心として、実証主義的な研究への志向が強まり、コーパス言語学の方法論が定着した。辞典編纂の期間に行われた約160回余りの共同研究会でなされた活発な討論、当時国内唯一の辞書学の専門学術誌であった『辞典編纂学研究』における論文の発表などは、その相当数が若い研究者の博士学位論文として実を結んだ。その主なものは『マルムンチ基盤国語研究叢書』(韓国文化社)として刊行されている。

- ・辞典式テキスト分析的な国語助辞の研究（イ ヒジャほか，1998年）
- ・国語の動詞連結構成に関する研究（カン ヒョンファ，1998年）
- ・国語形容詞研究（ユ ヒョンギョン，1998年）
- ・国語様態用言構文研究（キム チウン，1998年）
- ・辞典式テキスト分析的な国語語尾の研究（イ ヒジャほか，1999年）
- ・現代韓国語自動詞研究（ハン ソンファ，2000年）
- ・国語名詞の研究（チョン ヒジョン，2000年）
- ・韓国語教育と学習辞典（ソ サンギユ編，2003年）
- ・韓国語口語研究Ⅰ（ソ サンギユ・ク ヒョンジョン，2002年）
- ・韓国語口語研究Ⅱ（ソ サンギユ・ク ヒョンジョン，2005年）
- ・国語の助詞と意味域（イ ソンヒ，2004）
- ・現代国語「이다 (ida)」構文の研究（ナム キリム，2004年）
- ・現代韓国語助詞の研究Ⅰ（コ ソクジュ，2004年）

これらの研究者は、いずれも国語学の理論を身につけた言語学者としての資質と、語彙分析や辞書編集などの実際的な経験を積んだ辞典編纂の専門家としての能力とを兼ね備えた、貴重な人材である。

また、コーパスの電算処理、辞典原稿の電算管理及び構造解析にかかわるプログラム開発、言語情報の電算処理などに必要な専門家を養成するために、1997年に大学院に「国語情報学」課程（修士・博士課程）を設置し、体系的かつ実際的な教育プログラムを用意した。さらに、2000年からは、一般人のための講習会としての性格を有する「国語情報化アカデミー」を毎年夏に開催している。

6. おわりに

本論では、80年代中旬以降に行われた ILIS の研究活動を、大きく三つの側面から紹介した。特に、大規模言語資料の収集と活用、そしてそれを利用した辞典の編纂が、それ自体で完結したのではなく、隣接領域間の研究交流の活性化、ならびに実証主義的な言語研究の定着に大きく寄与してきたことを強調した。

辞典はすべての言語学的研究の統合点と言われる。辞典の編纂を通じて蓄積された方法論と基礎言語資料を活用して、韓国語文法の総合的記述を完成させることが今後の課題である。

ILIS の所在地・連絡先は以下のとおりである。

韓国ソウル市西大門区新村洞134 延世大学校 言語情報研究院

Tel. (02) 2123-3513 Fax. (02) 393-5001

また、ILIS に関する情報は以下の URL から得ることができる。

<http://lex.yonsei.ac.kr>（韓国語）

参考文献

(原文はすべて韓国語)

- キム ハンセム・ソ サンギユ (1999) 「延世コーパス I の構想と実際」『言語情報の探求 1』, 延世大言語情報開発研究院
- キム ヒョンジョン (2004) 「韓国語口語コーパスの構築と活用」, 朝鮮語研究会編『朝鮮語研究』 2, 163-191, 東京:くろしお出版
- キム フンギユ・カン ボムモ (1996) 「高麗大学校韓国語コーパス 1 (KOREA-1 CORPUS): 設計と構成」『韓国語学』 3, 韓国語学会
- キム フンギユほか (1998, 1999, 2000, 2001) 『21世紀セジョン計画国語基礎資料構築 (最終報告書)』, 文化観光部
- ナム ユンジン (1999) 「均衡コーパス構築のための実験的研究 (1)」『言語情報の探求 1』, 延世大言語情報開発研究院
- ソ サンギユ (1998a) 『延世コーパス 1-9 を対象にした現代韓国語の語彙頻度 (上・下)』, 延世大学校言語情報開発研究院内部報告書 (CLID-WP-98-02-28)
- ソ サンギユ (1998b) 「コーパス (corpus) 分析に基盤を置いた韓国語辞典の編纂」, 第49回朝鮮学会大会発表論文, 天理
- ソ サンギユ (1998c) 「コーパス分析に基盤を置いた単語頻度の調査とその応用—‘延世コーパス’を中心に—」『ハンゲル』 242, ハンゲル学会
- ソ サンギユ (1999) 『言語情報の探求 1』, 延世大学校言語情報開発研究院
- ソ サンギユ (2000, 2001) 『21世紀セジョン計画国語特殊資料構築 (最終報告書)』, 文化観光部
- ソ サンギユ (2001a) 「コーパスを利用した国語文法研究の現況と方向」, 高麗大学校民族文化研究院国語研究所編『21世紀国語情報化と国語研究』, 89-130
- ソ サンギユ (2001b) 「コーパスの註釈と韓国語基本語彙意味頻度辞典」『計量言語学』 1, 図書出版月印, 57-104
- ソ サンギユ (2002a) 「韓国語コーパスの構築と課題」『韓国語と情報化』, 太学社, 255-292
- ソ サンギユ (2002b) 「韓国語均衡コーパスの現況と課題」, パク ヨンスン編『21世紀国語学の現況と課題』, 韓国文化社, 149-173
- ソ サンギユ (2002c) 「国語情報学研究の現況と方向」『国語学研究50年』, 梨花女子大学韓国文化研究院, 431-463
- ソ サンギユ (2002d) 「韓国語基本語彙とコーパス分析」, パク ヨンスン編『21世紀韓国語教育学の現況と課題』, 韓国文化社, 361-396
- ソ サンギユ・ク ヒョンジョン共編 (2002) 『韓国語口語研究 (1) —口語転写コーパスとその活用—』韓国文化社
- ソ サンギユ・ク ヒョンジョン共編 (2005) 『韓国語口語研究 (2) —大学生口語転写コーパスを中心に—』, 韓国文化社
- ソ サンギユほか (1998) 「コーパス構築の方法論に関する研究」『21世紀セジョン計画国語基礎資料構築 (最終報告書)』, 文化観光部
- ソ サンギユほか (2003) 『外国人のための韓国語学習辞書』, 文化観光部・韓国語世界化財団
- ソ サンギユ・ハン ヨンギユン (1999) 『国語情報学入門』, 太学社
- ユン ジュンテ・チェ キソン (1999) 「KAIST コーパスに関する考察」, KAIST CS Technical Report Archives (CS-TR-99-139)

- イ サンソプ (1988) 「コーパス言語学的に見た辞書編纂の実際問題」『辞典編纂学研究』2, 塔出版社
- イ サンソプ (1990) 「単語頻度を推正するためのコーパス資料収集の実際」『辞典編纂学研究』3, 塔出版社
- イ サンソプ (1996) 「コーパス言語学の基本前提」『辞典編纂学研究』5・6, 塔出版社, 29-60
- チョン ヨンミ (1993) 「国語語彙の統計的特性とその応用」『人文科学』69・70, 延世大文科大学
- チョン チャンソプ (1989) 「韓国語語彙コーパスの標本選定基準」『ハングル及び韓国語情報処理学術発表論文集』
- チョン チャンソプほか (1990) 「国語単語頻度の調査標本の選定基準」『辞典編纂学研究』3, 塔出版社, 7-69