『インターネット技術を活用したマルチリンガル言語運用教育システムと教育手法の研究』，
基盤(B)(2) 課題番号 14310218，東京外国語大学，2005，379-390.

ユビキタス環境

佐野　洋
東京外国語大学外国語学部教授

1.　はじめに

いつでもどこでも，ネットワーク，コンピュータやコンテンツ等を自在に意識せずに利用できるユビキタスネットワークの実現に向けて研究開発が進められている。平成 15 年度からは総務省が委託研究によって「ユビキタスネットワーク技術の研究開発」を進めている。ユビキタスネットワークを支える主要技術には，(1) 超小型チップネットワーキング技術，(2) ユビキタスネットワーク認証・エージェント技術，(3) ユビキタスネットワーク制御・管理技術に関する研究開発などがある。

言語教育におけるユビキタス環境は，ユビキタスネットワークを実現する主要技術基盤を背景とした，いつでもどこでも利用できる多言語の語学学習環境といえる。こうした語学学習環境を実現するには，情報技術を利用する語学学習の方法論の研究，学習アプリケーションへのモバイル技術への適用の研究，学習コンテンツの多言語教育への拡張やメディアの拡充のための教材の収集と記録方式の研究，外国語教育ノウハウのデータベース化と教育内容のモジュール化(学習内容の構造化)のための課題の検討，学習教材の自動生成・自動変換に関する研究が必要となる。

本章は，言語教育におけるユビキタス環境を指向した語学学習教材の作成の方法論を述べる。この方法論は，コーパス利用技術を用いた教材作成の効率化と作成教材コンテンツに関するものである。

1.1.　ユビキタス環境と語学教材開発

社会基盤として情報環境が整備され，情報ネットワークを活用した教育(e-Learning)環境が充実しつつある。e-Learning は，ユビキタス環境における語学教育を実現する可能性を有し，語学教育に新しい可能性を開くものでもある。しかし，そのコンテンツは，既存の紙媒体用の教材を電子化した教材も多く，コンピュータの持つ機能を十分に活かしていなかったり，e-Learning システムを，省力化の目途に自主学習に利用したりするなど，必ずしも既存の教育枠組みと適切に組み合わされていない。そのため，ユビキタス環境が技術的に実現したとしても，学習コンテンツが従来のそれと変わらないことから，ユビキタス環境が提供する，いつでもどこでも利用できる語学学習の機会を十分に活かしきれない。

1.1.1.　教育枠組み

社会基盤が未整備であった頃，教育を広く行き渡らせるため，学習工程は，規格化(カリ

キュラム，教科)，分業化(学校階段，科目別)，同時化(修学年齢)，集中化(学校)によって効率化された。高度情報化が進展する現在，ネットワークが社会基盤となって，同時化と集中化の制限が大幅に緩和された。e-Learning システムなどの情報技術は，いわゆる学習場所と学習時間の制限をなくしている。教授法の発達によって分業化の制限も次第に緩やかになってきている。ユビキタス環境が実現すると，こうした制限はさらに緩和される。

1.1.2. 多用な教材ニーズと教材作成効率

学習行為の分散化(必要な事柄を必要な時に学習すること)が進んでいる現在，学習内容の制約の解消，すなわち規格化された教科から多様な教科の実現が求められている。例えば，ESP(English for Specific Purposes)もその一つで，従来の規格化された一般英語学習の枠組みでは教授できない。ビジネスを取り巻く環境のボーダーレス化，雇用の流動化や業務の国際化への変化に対応するため，専門的な分野で活躍できる英語力を持つ人材が強く求められている。このような専門分野の英語力向上を目指した英語学習(ESP)は，英語教育分野でもその必要性が認識されつつある。企業内教育においても ESP 適合の教育教材が学習者に与えられることが望ましい。

一方，教材開発には依然，労働集約的な作業で，その教材作成コストの削減の取り組みがとりわけ必要である。インターネット上を流通する膨大な電子化テキストと自然言語処理技術を応用することで，専門分野に合わせた教育教材(ESP 適合の教材コンテンツ)を，その作成コストを低減しながら，効率良く作り出すことが求められている。

1.2. 教材作成とコーパス利用

言語には，伝達機能だけでなく，思考の道具としての機能がある。我々は，ことばによって思考をしている。自らが言いたい事柄を思考によって明確に認識し，考えをまとめ，そして表現する。一方で，相手が言った事柄や記述されている事柄の中にある考えを整理し，その内容を理解する。こうした行為は，ことばの基本的な運用能力である。思考力なしにことばをうまく機能させることはできない。

思考とは何であるかを探求することは極めて困難な事柄である。思考への直接的な接近は難しいものの，思考の断片をことばの表現に求めて，その特徴を調べることは可能なアプローチである。そして，それに呼応するように，語学教育でも，思考伝達の直裁的な伝達方法を教えることが理想であるにも関わらず，教育教材は，思考の断片を表現する表現形式を羅列することによって構成されることが多い。

英語母語話者でない日本人が英語を教授する場合，その教材作成には母語話者の介在があるのが通例である。一方，日本語を母語とする日本語教育者はどうか。日本語を専門に教授しているとはいえ，その読書範囲や生活空間によって母語者としての内省によって得ることの多い学習用例の抽出には偏りがあることも否めない。

そこで本章では，語学教育素材開発のためのコーパス利用の方法論を説明する。これは

コンピュータを用いる言語分析ツールを利用することを前提とし，広範な表現の分布を高速に検索して，適切な表現例を見つけだすことを意図したものである。言語分析ツールを利用した教育素材を利用することで，いわゆる母語話者の言語直感に頼るのではなく，言語の運用事実に基づいた用例を教育教材に反映することができる。

1.2.1. 学習プロセス

学習は，プロセスの視点から捉えると，学習者にとっての価値を生み出す活動の集合体と考えられる。効果的な学習行為が学習プロセスには求められ，各活動は価値創造のために効率化の仕組みがある。未だ社会資本や社会基盤が整備されていない頃，学習プロセスを構成する各活動は，表 1 に示すように各種の制限を与えることで効率の向上を図ってきた。

表 1　学習プロセスと効率化にともなう制限

規格化	学習内容の制限……カリキュラム
分業化	学習領域と範囲の制限……学校階段，科目別
同時化	学習機会の制限……修学年齢
集中化	学習場所の制限……学校

例えば，社会基盤である交通網が発達することで，学習者は集中化の制約から一部解放されてきた。放送大学は，同時化の制約を緩和し，同時に放送網技術や衛星通信技術を利用することで集中化の制限も緩和した。

高度情報化社会の現在，情報通信技術が社会基盤として整備されつつある。学習プロセスの効率化のために設けられていた制約は一層緩和されてきている。とりわけインターネットなどネットワーク網の発達と通信帯域の広がり，ユビキタス環境の整備に伴う，ネットワークを通じた自己学習支援環境(e-Learning)の発展は，個々人のレベルにおいても，同時化と集中化の制約を大幅に緩和している。学習場所と学習時間の制約が緩和された結果，いつでもどこでも学べる環境が整ってきた。生涯教育の推進など学習者要求の多様化から分業化もその制約を緩和せざるを得なくなっている。IT 技術，環境技術や遺伝子技術などは，新しい学問分野を形成し，学習領域が新規に生まれている。

1.2.2. 学習効果の改善

学習プロセスの改善のため残されている重要な課題は，学習に費やす時間の短縮である。効率的な学習行為を支援することだ。表 1 における規格化の制約の緩和に関する。集中化に連動した多数の学習者向け教材から，多様性のある(個人適合の)教材の開発が目標となる。

学習教材は学習法とも深い関係を持つ。学習法には，教育の方法論に指針を与える枠組み(アプローチ)，学習内容を決める教授設計(インストラクション・デザイン)，そして，教

育現場で学習者の意欲を左右する技法(テクニック)があって，それらを実現する行為は，いずれも極めて属人性が高いことが特徴である。すなわち学習法に関わる行為は，労働集約作業であって，直裁的な技術の応用や適用では解決が困難な側面も多分にある。

筆者は，教授設計の中の教材作成行為に焦点をあてる。教育素材は，数多くの資料にあたり，その資料から学習段階に適切であると考えられる素材を抽出することで得る。この作業は高度に知的な検索過程として抽象化できる可能性がある。そしてこの検索作業は，コーパスと情報技術で置き換えることができるから，知的作業部分を支援する仕組みを工夫することで，教育素材の作成のための支援環境を開発できる可能性がある。

1.2.3. 学習課題の種類

学習課題の種類は，[鈴木 02]によると(1) 認知領域の課題，(2) 運動領域の課題，(3) 情意領域の課題の 3 つに区分できるという。教材作成の際にも三領域のいずれに分類できるかを確認することが肝要であることを指摘している。

運動領域の課題は，学習行為に運動技能の訓練を伴うことが特徴で，従って，その教材は，運動を促進するための補助的な側面が強い。情意領域の課題は，人に接する気持ちや行動規範に関わるもので，その教育が難しいのと同様に，教材作成も困難であると考えられる。

それに対して認知領域の学習課題は，計算，物事の暗記や規則の適用能力に関わるもので，この学習課題に関連する教材の多くは,資料検索の高速化によって作成効率の向上を見込めるものである。本章で示す語学教材作成のためのコーパス利用法は，この認知領域の学習課題に関する教材素材の作成を支援する。その直接的な成果は，個人のニーズ，学習段階や学習スタイルに適う学習コンテンツとして具体化できることにある。

なお，認知領域の学習課題は，事柄を覚えて思い出す記憶課題と規則の運用能力の課題があってそれぞれ，言語情報と知的技能として明確に区分することが教材作成でも必要であることが指摘されている。

2. 対照言語学の視点を考慮した多言語コーパスの作成とその利用(1) 日本語

2.1. はじめに

効果的な言語分析を行うには，言語体系の正確な知識が欠かせない。言語の知識は，運用実態である文や文章を注意深く観察し分析することで得られる。文の記述する内容を心の中で広げてみると，そこには必ず叙述状況がある。叙述状況の中の関与者に注意を向けると，その関与者は性質を表し，何らかの状況に埋め込まれていることに気づく。関与者のない状況，状況に関わらない関与者だけの認知はない。

しかしながら，言語は，写真を使った状況描写に見られるような均質な状況叙述を行おうとはせず，ある種の関与者や特定の関係だけの叙述を断然好む偏りがある。さらに偏りを具現化する手段も多様である。[4]によると，例えば，文法関係を取り上げてみると，も

っぱら語順で示そうとする英語のような言語もあるし，ロシア語のように，その多くを形態法にゆだねる言語もある。前者は，文法関係が語用論的な役割から独立し，後者は，語用論的な役割によって語順が決定するという。さらに意味役割と文法関係の相互作用は，英語に比べ，ロシア語に顕著に現れるという。

2.1.1. 対照言語学的な視点

ことばを取り巻く諸々の現象は，不均質に顕在化すること，また，具現化の手段が多様であることがわかる。調査しようとする項目(文法関係，意味役割，語用論的機能)の手がかりが，単一の言語内で程良く現れることは保証されない。このことは，複数言語に対して，個々の言語が好む偏り部分を特定し，且つ，当該部分を集中的に調査すれば，効果的に言語の知識を得ることができるだろうことを示唆している。

例えば，日本語のアスペクト研究は，ロシア語のアスペクト研究の影響を強く受けているという。[5]で説明されるようにロシア語の動詞はアスペクトの範疇を持っている。すべての動詞は，動作を全体的で限界のあるものとして表現するか否か(完了体，不完了体)を形態上で区分する。

また，英語は文法関係によって決定されるかなり固定した語順を持った言語であり，しかも，多くの統語的プロセスは，線形順序の変化という点から記述できる[5]から，例えば，従属節の主語や目的語が音形をなくし，主節の目的語になる現象などは，英語の分析研究の影響を受けて日本語でも応用研究された。

2.1.2. 多言語コーパス

単一の言語についての包括的な体系知識は，言語間の違いの調査と分析によって得られるのである。多言語コーパスは，その設計と収集指針に対照言語学的な視点を明示的に反映することが重要である。単一の言語，例えば，日本語コーパス作成についても，日本語が世界の諸言語とどのような点で類似し，どのような点で異なるかを認識した上で作成するべきだろう。

本章では，対照言語学的な視点に基づく多言語コーパス作成の設計指針について述べる。その指針の中では，コーパス収集を効果的なものとする基準を示し，その基準を設ける背景となった言語運用の分析モデルを示す。また，多言語コーパス作成のための言語調査ツールについても説明する。

以下では，言語運用の分析モデルと多言語コーパス作成のための検索ツールを説明する。対照言語学的な視点に基づく多言語コーパス作成の設計指針は，次章で述べる。別稿では，統語範疇が豊富で形態法が発達しているロシア語について概観し，タグセットの設定の試みについても説明する。同時に幾つかの用例を挙げて，問題点も指摘する。

2.2. 対照言語学的な視点によるコーパス収集

2.2.1. コーパス収集のための言語の分析モデル

ことばの運用実態の中で，語の性質や意味を見いだそうとするアプローチは，コーパス言語学として一分野を形成するほどに成果を上げている[6,7]。このアプローチでは，文中の形の変化を，コンピュータを使って把握し文の分析をする。我々の形の変化の分析モデルは[pp.87～pp.89, 8]に従っている。このモデルは，語は語連鎖としての句の外形形態と意味との関連性を観察することを目的とし，表 1 のようにまとめられる。

表 2　Model of Extended Lexical Units

	RELATION	Constituent
(1)	COLLOCATION	Collocate
(2)	COLLIGATION	grammatical category
(3)	SEMANTIC PREFERENCE	lexical set
(4)	DISCOURCE PROSODY	descriptor of speaker attitude and discourse function

(1)から(4)に従って抽象度があがる。(1)は，表層の形態表現もしくは単語の連鎖表現で文中に直接観察できる。(2)は，間接的な観察である。具体的な形ではなく抽象化した文法用語を利用して文の形を観察する。(3)は，意味素性を応用する。文法用語よりもさらに抽象化が進んだ存在論的な論理関係(階層関係，全体-部分の関係，因果関係，前後関係，対立関係など)を利用する。(4)は，発話状況における世界知識を利用しなければ分析できない。

2.2.2. コーパス収集の方針と基準

本章でのコーパス収集における関心は，ある言語の言語体系を網羅したり，文法現象を列挙したりするという問題に，その言語を表層レベルから始まり，談話レベルまで深く子細に調査分析を進めるという一般的な解決法で迫るのではなく，形態上の特徴に的を絞ることにある。但し，複数の言語に視野を広げ，特定の言語における特徴現象の顕在化の不均質さを補償する。前節で示した分析モデルの(1)もしくは, 高々(2)の部分までを利用する。仮に，ある言語で(3)や(4)のレベルの現象を観察するときには，他の言語の(1)や(2)の結果を利用するのである。換言すると，その概念が内包で示される抽象化に依存する部分をできるだけなくし，代わって外延で具現化される現象を収集し，分類基準として扱うアプローチである。

2.2.3. 収集粒度の平準化

ある言語現象に対して形態的な特徴や対立からコーパスを収集し，言語構造を他の特徴(文法関係，意味役割，語用論的機能)と相関させて有効な説明を見いだす。それら説明を代表する用例を他の言語に翻訳した対訳コーパスを用意する。ある現象について形態的な特

徴や対立が表れない言語では，対訳用例を基準に言語構造と他の特徴を関係づける。この収集と分類の手法の主な利点は，多言語間で分析粒度の平準化を図っている点にある。

ここでの作業前提は，ある言語で，少なくとも‘“断然好む偏り”が見られるある領域に関しては，収集用例の総体(コーパス)は，そこから言語構造を他の特徴と相関させて有効な説明を見いだすのに十分な幅をもった言語母集団になっているということである。こうして単一の言語分析に執着していたら起きたかもしれない，ある種の偏りをさけることができる。

ある単一の言語ないしは，単一のグループの言語に過度に偏ることのないようなコーパスの収集が大切である。次章では，形態法の発達したロシア語について外観し，コーパスのタグセットについて議論する。

2.3. 言語調査ツール

2.3.1. 分析の広さと深さ

単一の言語について，(1)から(4)へと抽象度を上げて言語構造と他の特徴の関係をより深く研究していく立場がある。それに対して我々の立場は，複数の言語について，さほど抽象的なレベルを設けずに，形態特徴が顕著に表れる表示レベルで，言語構造と他の特徴を広く浅く探求するものである。ここでは用例収集のための言語調査ツールについて述べる。

2.3.2. 軽量ツールを見直す

2.1. 節で示した分析モデルの(1)や(2)の調査に利用するソフトウェアは，既に十何年も前に開発されている。解析技術的な革新性はないが現代的な OS 上で安定して機能するし，文系研究者にも使い易い。さらに計算能力が向上したハードウェアを利用することで，高速実行も期待できる。

2.3.3. 調査ツール

多言語コーパスを作成する上で，各言語の調査ツールの整備は欠かせない作業である。もっか，日本語，英語，ロシア語，スペイン語の各言語調査ツールを整備している。

英語については，フリーウェアの言語調査ツールも多いほか，商用化された言語調査ツールを比較的低価格で入手することができる。

日本語については，筆者の佐野は，既存の日本語研究用ソフトウェア[4]を，適応保守 の工程を経て開発した[1,2]。AWK 言語で記述された日本語研究用プログラムを，処理環境の変化に適用させる保守を実施し，Perl 言語で再実装した。同時に現代的なユーザーインタフェース部分を新規に作成した。書き換えられたソフトウェアは，機能面で問題がなく，現代的な OS に適合している。また，Windows XP(R)のように OS レベルで Unicode 対応になっていること，Perl 処理系の版が 5.8. に向上し Unicode 文字列を扱えるようになったこと

から，筆者は現在，適応保守で再開発した日本語調査用ソフトウェアの多言語化を進めている。

なお，ロシア語調査ツールとスペイン語調査ツールは別稿[5]を参照されたい。

2.4. おわりに

本章では，いわゆる形態的類型論の考え方を応用したコーパスの設計指針と具体化の手段について述べた。現在，こうした各言語用の言語調査ツールを利用して，用例の収集を始めている。

2.2. 節で示した分析モデルの COLLOCATION を重視する用例の収集は，(1) 収集に利用する言語分析ツールも軽量でよく複雑な仕組みを利用しないのでコストが安い，(2) 顕在化した形態を観察することから抽象概念の入り込む余地が小さく，言語現象の解釈が安定するなどの利点がある。ある言語で断然好む偏りを効率よく見つけること，収集した用例集を正確な翻訳をともなって多言語化することが今後の課題である。

3. 対照言語学の視点を考慮した多言語コーパスの作成とその利用(2) ロシア語

3.1. はじめに

本章では，対照言語学的な視点に基づく多言語コーパス作成の設計指針について述べる。以下では，統語範疇が豊富で形態法が発達しているロシア語について概観し，多言語コーパスのタグセットの設定の試みについて説明する。同時に幾つかの用例を挙げて，問題点も指摘する。

前章では，コーパス収集を効果的なものとする基準を示し，その基準の背景となった言語運用の分析モデルを説明した。また，多言語コーパス作成のための言語調査ツールについても説明した。

3.1.1. 対照言語学的な視点

前章で示したように，言語は，写真を使った状況描写に見られるような均質な状況叙述を行おうとはせず，ある種の関与者や特定の関係だけの叙述を断然好む偏りがある。さらに偏りを具現化する手段も多様である。そして，さらに，ことばを取り巻く諸々の現象は，不均質に顕在化すること，また，具現化の手段が多様であることから，調査しようとする項目(文法関係，意味役割，語用論的機能)の手がかりが，単一の言語内で程良く現れることは保証されない。このことは，複数言語に対して，個々の言語が好む偏り部分を特定し，且つ，当該部分を集中的に調査すれば，効果的に言語の知識を得ることができるであろう。

3.1.2. 多言語コーパス

前章で示したように，単一の言語についての包括的な体系知識は，言語間の違いの調査と分析によって得られる。多言語コーパスは，その設計と収集指針に対照言語学的な視点を明示的に反映することが重要である。単一の言語，例えば，日本語コーパス作成についても，日本語が世界の諸言語とどのような点で類似し，どのような点で異なるかを認識した上で作成するべきだろう。

例えば[5]は，英語は，いろいろな面で，世界の言語全体の中で決して典型的でもなく，また，多くの違ったタイプの中のひとつを代表しているにすぎない。従ってまた，もっぱら英語の分析だけに目を向けた言語理論は，このような要因に毒される危険性があるだろうと指摘する。

以下では，対照言語学的な視点に基づく多言語コーパス作成の設計指針を述べる。まず，形態法が発達しているロシア語を概観する。いわゆる形態的類型論の考え方を応用したコーパスの設計指針を示すため，[5]を参照してタグセットの検討を行う。幾つかの具体例を挙げて問題点の指摘をする。ロシア語とスペイン語の言語調査ツールについて簡単に説明する。

3.2. 対照言語学的な視点によるコーパス収集

多言語コーパスを作成する際の問題の一つは，どのようなタグの集合を妥当とすべきか，という点である。豊かな形態法を有するロシア語の統語範疇を，最大範疇基準とするタグセットを考えることは，魅力的な指針に思われる。

3.2.1. ロシア語の統語範疇の概観

[3,6,7]によると，ロシア語は，インド・ヨーロッパ語族の言語で，その中のスラヴ語派の一つ，東スラヴ語に属する。ロシア語は典型的な屈折語である。性，格，数を表示する複雑な語尾交代の体系を持つ，きわめて形態法が発達している言語である。

以下，参照文献によると，例えば，代名詞を含めた名詞類は格変化し，性，数による変化のほか，主格，生格，与各，対格，造格，前置格の 6 つの格を区別する。主格と対格は，用法と意味の対応関係が比較的安定している。

また，動詞は，アスペクト，ボイス，法，時制，人称，数などの統語範疇が形態的手段で具現化されている。アスペクトの範疇について，他のスラヴ諸語と同様に全ての動詞は，動作の全体的な把握を特徴とする完了体か，あるいはその特徴を欠いている不完了体のいずれかに所属する。ボイスは，能動態と受動態を区別する。法の範疇では，不定形と分詞を除いて，直説法，命令法，仮定法の 3 つの区別があるという。直説法は，時制の範疇を持ち，現在形，過去形，未来形がある。命令法では，一人称命令法と二人称命令法が形態上区別される。前者は，共同作業への促しとしての意味があり，後者は，相手に対する命

令としての意味を持つ。仮定法は，動詞の過去形相当の形態に特定の助詞を合わせて表現する。分詞は，副動詞(副詞的機能)と形動詞(形容詞的機能)の区別がある。前者は，完了体と不完了体で形の区別がある。後者は，能動体と受動態の区別とさらにそれぞれ現在形と過去形がある。なお，語順は相対的に自由である。

このようにロシア語では，文法関係を形態法で標示する。この統語範疇のタグ集合としての妥当性にさらに検討を加える。

3.2.2. 品詞分類と文法関係

名詞に性，数区分のない言語が，名詞に性，数区分が認められる言語に比べて異常であるという根拠はないが，動詞に完了体と不完了体の区分を設ける有用性はある。動詞の体(アスペクト)区分は，複文(もしくは従属節)の用法と関係し，存在論的な論理関係の因果関係や前後関係と相関し，意味役割に関係している。

「溺れたけれど助かった」などの用例は，動詞の完了体と不完了体の区分を要求する。主語や目的語などの文法関係について，[5]は，「特定の言語に特定の文法関係が存在するというためには，その主張が言語内的 にも言語間的にも十分に正当化されかねればならない」と指摘する。例示の中で，英語の間接目的語は，(おそらくは)文法関係ではないという。6 つの格を区別する範疇によって名詞の形態特徴と文法関係を相関させることができる。しかしながら，形態格と文法関係の間には，ある程度の不一致があることも指摘されている。形態格名をタグ名として利用する際には，この不一致の程度を確認することと，それを他の言語の形態格で補間することが必要である。

3.2.3. 意味役割

ロシア語の活動体名詞は，有生性を形態的に示す。文法関係が意味役割と緩い相関しかない英語などに比べると有用な標示である。しかしながら，この有生性の性質は，[5]によると，働きかけの能力(制御力)として，ある言語では，使役を表現する構文に違いを生じさせたり，別の言語では，経験者と被動者を区別する形態格(能格)として顕在化したりする。さらに，例えば，英語では，“I fell”と言った場合，“I”の制御力の程度は何も示されていないが，“I”が故意に倒れたのか，あるいは，不注意で倒れたのかを文法的に(形態格で)示す言語もあるという。

ロシア語の統語範疇の枠組みだけでは，意味役割の重要性を見失う可能性がある。コーパスに付与するタグによって文法関係を通じた意味解釈を実現するには，範疇名の拡張が必要である。

3.2.4. 語用論的役割

[5]によると「語用論的ないし談話的役割とは，本質的に同じ情報ないし，同じ意味内容

が違った構成をとることによって情報の新旧の流れを反映するような，さまざまなあり方のことである」という。そして，事柄の定性/不定性と話題/焦点が語用論的役割を表現する用語として提案している。また，[5]は，英語には，一般に題目と焦点に対する文法化がないことを指摘し，ロシア語ではそれを語順で示すという(文頭が題目で文末が焦点)。題目が形態法で示される日本語や朝鮮語の用例で補間することで，タグセットに題目を加えることができる。さらに日本語は話者の意図表現が豊かである。例えば，「の」は，その形態によって事柄の事実性(定性)や説明的なモダリティ(焦点)を標示する。以下は「の」の広範な使用の様子を示す。

表 3　日本語の「の」の朝鮮語翻訳例

日本語	朝鮮語訳	
切り抜いたのだろうか,	오려 낸 것일까?	○
つづいているのだが,	계속되고 있지만,	×
無視していたのだ。	무시하고 있었던 것이다。	○
答えたのだ。	대답했다。	×
嘘だったのだな, (と気付いたが)	거짓말이라는 걸 (눈치 챘지만)	×
信じたかったのだ。	믿고 싶었던 것이다。	○

表右端のチェック欄は，「のだ」が翻訳に反映されているかどうかを示す。朝鮮語にも「の」に相当する準体の接辞があるが，翻訳例で見るように日本語は，その使用範囲が広い。日本語や朝鮮語用例によって，モダリティタグなども考察の範囲に入る。

3.3.　言語調査ツール

英語と日本語の調査ツールについては，前章で示した。

3.3.1.　ロシア語タグ付けソフトウェア

もっか利用しているロシア語のタグ付けソフトウェアは，AOT(URL：www。aot。ru)から提供されているツールである。AOT は自動テキスト処理プロジェクトを推進する組織である。ツールは，研究目的に限り利用できる。英露機械翻訳プロジェクトの DIALING システムの副産物のようである。動作環境は，ロシア語版 Linux / Black Cat(Red Hat7。2 に対応)でソースコードをコンパイル(GCC++)して利用する。なお，Windows 版もある。

3.3.2.　スペイン語タグ付けソフトウェア

Universitat Politecnica de Catalunya から提供されている PC-MSTools を利用している。研

究目的で利用することができる。例えば，“Sr. Mesa se mesa la barba al lado de la mesa.” は，次のようにタグ付けされる。

El	el	DA0MS0
Sr.	sr.	NC00000
Mesa	mesa	NCFS000
se	se	P0000000
mesa	mesa	NCFS000

…..

3.4. おわりに

本章は，対照言語学的な視点に基づく多言語コーパス作成の設計指針について述べた。統語範疇が豊富で形態法が発達しているロシア語について概観し，多言語コーパスのタグセットの設定の試みについて説明した。同時に幾つかの用例を挙げて，問題点も指摘した。今後，ロシア語の統語範疇をもとに拡張を行い，多言語コーパス用のタグセットを確立する予定である。

参考文献

[1] 佐野洋，幸松英恵,「ソフトウェア再利用による語彙調査用ツールの開発」,2003 年，言語処理学会講演論文集。

[2] 佐野洋，「ソフトウェア再利用による日本語研究のための分析ツールの開発」,2003 年,電子通信学会全国大会講演論文集。

[3] 城田俊,「現代ロシア語文法 中上級編」,東洋書店,2003 年。

[4] 中野洋,「パソコンによる日本語研究法入門」,笠間書院,1996 年。

[5] バーナード・コムリ著，松本克己，山本秀樹訳,「言語普遍性と言語類型論」，ひつじ書房, 1992 年。

[6] Michael P. Oakes, “Statistics for Corpus Linguistics”, Edinburgh University Press, 1998.

[7] Michael Stubbs, “Words and Phrases Corpus Studies of Lexical Semantics”, Blackwell Publishers Inc., 2001.