

多言語平行コーパスのための 「言語学におもしろい100の文」

100 linguistically interesting sentences for a multilingual parallel corpus

山 崎 直 樹
YAMAZAKI Naoki

The Multilingual Processing Project, launched at Osaka University of Foreign Studies (present Faculty of Foreign Studies, Osaka University), is now trying to build various types of multilingual parallel corpora.

As part of this project, I selected almost 100 sentences that seemed to be linguistically interesting. In this context, “linguistically interesting” means that the cross-linguistic data associated with these sentences can make some contribution to the study of comparative and typological linguistics.

In this paper, I'll introduce the Japanese version of these 100 sentences and comment on why they are “interesting.” The sentences include copulative sentences, high/low transitivity sentences, double nominative constructions, double objective constructions, passive/causative/middle voice sentences, action-result constructions, comparative sentences, existential sentences, various types of interrogative sentences, expressions of semantic roles such as instrumental/goal/route/purpose, complicated noun phrases, variations of relative clauses and others.

Keywords

- 1) multilingual parallel corpus, 2) cross-linguistic data, 3) linguistic typology,
- 4) comparative linguistics, 5) linguistically interesting

キーワード

- 1) 多言語平行コーパス, 2) 言語横断的データ, 3) 言語類型論, 4) 対照言語学,
- 5) 「言語学におもしろい」

0. 序

大阪外国語大学（現在は大阪大学外国語学部）で発足した多言語処理プロジェクトは、その活動の一環として多言語平行コーパスの構築を試みている。プロジェクトのメンバーの1人である筆者は、このコーパスに収録するデータの一種として、「言語学的におもしろい100の文」を選定した。「おもしろい」とは、この「100の文」を種々の言語に翻訳し対照させたデータは、統語構造の対照言語学的研究や言語類型論的研究にとって興味深いデータを提供できる（と思われる）という意味である。

以下、この100文（実際は約100文だが）の日本語版を紹介し、その文を選んだポイント（どこが言語学的におもしろいか）を簡単に解説する。

1. 資料

1.1 「等価」を表す文（「AはBである」タイプの文）

#001：わたしは教師です。

これは典型的なコピュラ文である。この例は、恒常的な時制／非限定的なアスペクトを持つ文であり、言語によっては、「コピュラ無し文」（＝2つの名詞句が何の標識もなしに連続するだけの文）になる可能性がある。

#002：わたしの母は弁護士でした。

典型的なコピュラ文だが、完了的な時制／アスペクトを持っている。#001で「コピュラ無し」になる言語でも、この場合には、コピュラが現れることが多い。また、コピュラのような状態性の強い述語は、文法化された時制／アスペクト辞と共起しない言語もある。

#003：わたしは、将来、医者になりたい。

「～である」と「～になる」を別の語彙で表すかどうか。

#004：今日は2月29日です。

「AはBだ」の述語部分が暦時表現の場合。このような述語では「コピュラ無し文」が現れることが多い。

#005：わたしの夫は30歳です。

「AはBだ」の述語部分が数量表現の場合。このような述語では「コピュラ無し文」が現れることが多い。

1.2 「描写」をする文(「Aは～だ」タイプの文)

#006: この橋は古い。

描写的／状態的な述語をもつ文。ふつうは形容詞述語文になると思われるが、形容詞という品詞をもたない言語ではどうなるか。

#007: この壁はかつては青かった。

描写的／状態的な述語をもつ文。状態的な述語と時制辞がどのように共存するか(あるいは、しないか)。

1.3 状態を表す文

#008: 新しい首相はサルにとてもよく似ている。

状態性の高い述語をもつ文。もし述語が動詞になるのであれば、どのようなアスペクトで表現されるか。

#009: わたしは近所の人と仲が悪い。

状態性の高い述語をもつ文。述語が形容詞の場合、二項形容詞になるかどうか。

1.4 自動詞文

#010: コウモリは夜飛ぶ。

自動詞文。「法則／真理」を表す内容が、恒常的な時制／非限定的なアスペクトで表されるかどうか。

#011: 赤ん坊が笑った。

自動詞文。完了的な時制／アスペクトをもつ文。この時制／アスペクトの違いが、#010との違いをもたらすかどうか。

1.5 他動詞文

#012: ネコはネズミを追う。

他動詞文。「法則／真理」を表す内容が、恒常的な時制／非限定的なアスペクトで表されるかどうか。

#013: コーチは旗を一本立てた。

他動詞文。完了的な時制／アスペクトをもつ文。この時制／アスペクトの違いが、#012との違いをもたらすかどうか。なお、この文は定性の低い目的語をもつことに注意。

#014: 兵士はその家を破壊した。

他動詞文。完了的な時制／アスペクトをもつ文。定性の高い目的語をもつことに注意。この目的語の定性の違いが、#013との違いをもたらすかどうか。

1.6 自動詞と他動詞のペア

#015: 我々は勝った。

#016: 我々はライバルチームに勝った。

この「勝つ」という意味の語は、多くの言語で、自動詞・他動詞の両用法をもつ1つの動詞か、形態的に派生関係にある2つの動詞で表されることが多い。そして、どちらの場合も、外項対応型／対格型のふるまいをすることが多い。

#017: 城門が開いた。

#018: 市民たちは城門を開けた。

この「開く／開ける」という意味の語は、多くの言語で、自動詞・他動詞の両用法をもつ1つの動詞か、形態的に派生関係にある2つの動詞で表されることが多い。そして、どちらの場合も、内項対応型／能格型のふるまいをすることが多い。

1.7 Yes-No 疑問文

#019: あなたはスパイですか？

コピュラ文のYes-No疑問文。肯定文ではコピュラ無し文になる場合でも、疑問では述語動詞が現れることが多い。

#020: 明日は日曜日ですか？

述語が暦時表現の場合のYes-No疑問文。肯定文ではコピュラ無し文になる場合でも、疑問では述語動詞が現れることが多い。

#021: あなたは20歳ですか？

述語が数量表現の場合のYes-No疑問文。肯定文ではコピュラ無し文になる場合でも、疑問では述語動詞が現れることが多い。

#022: この料理はおいしいですか？

状態的な述語の場合のYes-No疑問文。

#023：ネコはネズミを追いかけるか？

他動詞文のYes-No疑問文。恒常的な時制／非限定的なアスペクトの場合。

1.8 Wh疑問文

#024：キリンは何を食べますか？

他動詞文のWh疑問文。恒常的な時制／非限定的なアスペクトの場合。

#025：これは誰の聖書ですか？

限定詞が疑問詞になっている疑問詞疑問文。

1.9 二重主語

#026：ゾウは鼻が長い。

題目と主語をもつ文（二重主語文、あるいは大主語・小主語をもつ文、とも言う）。題目と主語が[全体-部分]の関係になっている。当該の言語がこの構造をもつかどうか。この構造をもたない場合、別のどのような構造で表現するか。

#027：この樹は葉が大きく花が小さい。

題目と主語をもつ文。題目と主語が[全体-部分]の関係である。1つの題目を共有して主述構造が連鎖をなしている。[全体]を示す名詞句と、その[部分]を表す名詞句との間の結束性がどのように表示されているか（あるいは、表示されていないか）。

#028：プレゼントはダイヤがいい。

題目と主語をもつ文であるが、題目と主語の間に、[全体-部分]のような明確な関係がない。このような構造を許容するかどうか。

1.10 二重目的語

#029：わたしは彼に辞書を一冊送った。

「送る」という意味をもつ動詞は、多くの言語で二重目的語を取る。#029では、間接目的語の定性が高く、直接目的語の定性が低いことに注意。

#030：わたしはわたしが編纂した辞書のある小学生に送った。

「送る」という意味をもつ動詞は、多くの言語で二重目的語を取る。#030では、間接目的語の定性が低く、直接目的語の定性が高いことに注意。この違いが、文の構造に影響を与えるかどうか。

#031：みんなが彼を「臆病者」と呼ぶ。

「AをBと呼ぶ」という意味の動詞が二重目的語を取るかどうか。

#032：我々はよその国から来た人を「外国人」と呼ぶ。

「AをBと呼ぶ」という意味の動詞が二重目的語を取るかどうか。「Aを」が長いフレーズの場合かどうか。

#033：強盗は彼からお金を少しまきあげた。

マイナス方向の授与動詞 (= 略奪動詞?) をもつ文。間接目的語の定性が高く、直接目的語の定性が低い。

#034：どろぼうはそのナイフをどこかの店から盗んだらしい。

マイナス方向の授与動詞 (= 略奪動詞?) をもつ文。間接目的語の定性が低く、直接目的語の定性が高い。

1.11 要望

#035：早くクリスマスが来てほしい。

要望 (願望) を表す節を持つ文。

#036：わたしは君にこの件を調査してもらいたい。

要望を表す使役構文。

1.12 使役

#037：姉はわたしにほうれん草をむりやり食べさせた。

「強制」の意味をもつ使役文。

#038：看護師はその学生を立たせた。

「強制」の意味をもつ使役文。「立てた」(他動詞文)と「立たせた」(使役文)の区別がある言語では、どちらを使うか?

#039：希望者を海外に行かせよう。

「許可」の意味をもつ使役文。「強制」と区別するかどうか?

#040：わたしは子供に風邪をひかせてしまった。

「放任」の意味をもつ使役文。「強制」と区別するかどうか？

#041：わたしは床屋に髪を切らせた。

主語と利害関係を有する「部分」に対する使役。

#042：家族全員で父親を説得して入院させた。

語彙的な意味の中に使役的な意味をもつ動詞が使役構文を形成できるかどうか。

#043：政府は革命家たちを自由にした。

#044：出て行く前に部屋をきれいにしなさい。

「ある状態にする」という使役構文をとるか、形容詞を使役的な意味を持つ他動詞として用いるか。

1.13 受身

#045：うちのロバがトラックにはねられた。

典型的な「被害」の意味をもつ受身文。

#046：わたしは飼っているイヌに手をかまれた。

主語と利害関係を有する「部分」に対する「被害」の受身。このような受身文が存在するかどうか。

#047：わたしは二番目の妻にも逃げられた。

いわゆる間接受動文。いわゆる「迷惑」の受身文。このような受身文が存在するかどうか。

#048：この小説はたくさんの人に読まれた。

#049：あのビルは19世紀に建てられた。

無生物主語の中立的受身文(=「被害」の受身ではない)。

1.14 中動態

#050：このナイフはよく切れる。

#051：この本はよく売れる。

いわゆる中動態の文。描写や叙述の対象を主語にした文で、述語部分が、動詞+叙述要素になるか、あるいは受動態に変換されるか。

1.15 存在

#052: その机の上にペンが一本あるよ。

存在文。「場所」の定性が高く、存在物の定性が低い場合。一般の文と語順が異なるかどうか。

#053: この道の両側にはたくさんの高層ビルが建っている。

存在文。「場所」の定性が高く、存在物の定性が低い場合。いわゆる存在動詞ではなく、一般の動詞の場合。動詞がどのようなアスペクト形式になっているか。

1.16 所在

#054: 君のペンはその机の上にあるよ。

所在を表す文。無生物で定性の高い存在物の場合。生物の所在と無生物の所在を区別するかどうか。

#055: わたしの孫はここにいる。

所在を表す文。生物で定性の高い存在物の場合。生物の所在と無生物の所在を区別するかどうか。

1.17 出現

#056: むこうのほうから警官が1人駆けてきた。

出現を表す文。「場所」の定性が高く、存在物の定性が低い場合。一般の文と語順が異なるかどうか。

#057: さっき出て行った警官がまた戻ってきた。

出現を表す文。定性の高い存在物の場合。

#058: 雨が降り出した。

気象現象の出現を表す文。一般の文と語順が異なるかどうか。

#059: 雪が降っている。

気象状態を表す文。継続/持続のアスペクトをもち、定性の高い主語の場合。

#060: この屋敷はよくお化けが出る。

超自然現象の出現を表す文。一般の文と語順が異なるかどうか。

1.18 消失

#061: ウシ小屋からウシが一頭逃げた。

消失を表す文。「場所」の定性が高く、存在物の定性が低い場合。一般の文と語順が異なるかどうか。

#062: いちばん若いウシがウシ小屋から逃げた。

消失を表す文。定性の高い存在物の場合。

1.19 内訳

#063: 4つのグラスのうち2つが割れた。

部分集合にフォーカスをあてる文。自動詞の場合。

#064: 弟は5つのリンゴのうち3つを食べた。

部分集合にフォーカスをあてる文。他動詞の場合。

1.20 結果

#065: 彼女は夫の母をハンマーで殴り殺した。

動作とその結果を表す文。意図的な結果の場合。複合動詞で表現するのか、フレーズで表現するのか、使役の標識を必要とするのか、しないのか。

#066: 彼女は夫の父を思わずフライパンで殴って死なせてしまった。

動作とその結果を表す文。非意図的な結果の場合。複合動詞で表現するのか、フレーズで表現するのか、使役の標識を必要とするのか、しないのか。

1.21 未達成

#067: 壁のポスターを剥がした(剥がそうとした)が剥がれなかった。

#068: 黒板の字を消した(消そうとした)が消えなかった。

結果が未達成になった行為。「～した」(完了アスペクト) を使えるのか「～しようとした」(試行アスペクト) でないとだめなのか。

#069: ゴミ箱の蓋を開けた(開けようとした)が開かなかった。

結果が未達成になった行為。「～した」(完了アスペクト) を使えるのか「～しようとした」(試行アスペクト) でないとだめなのか。

1.22 着点

#070：彼はボールを100m先の的に当てた。

行為の着点を明示する項をもつ文。「着点」がどのような格標識／側置詞で表されるか。

#071：脱いだ靴はベッドの下に置いてくれ。

行為の着点を明示する項をもつ文。行為の結果が存続するタイプの動詞の場合。「着点」がどのような格標識／側置詞で表されるか。

#072：泳いで対岸まで行こう。

行為の着点を明示する項をもつ文。目標となるのが地理的な場所の場合。「着点」がどのような格標識／側置詞で表されるか。

#073：図書館は夜9時まで開いている。

行為の着点を明示する項をもつ文。目標となるのが時刻の場合。「着点」がどのような格標識／側置詞で表されるか。

#074：彼女はからださがガタガタになるまで働いた。

行為の着点を明示する項をもつ文。目標となるのが、ある状態の場合。「着点」がどのような格標識／側置詞で表されるか。

1.23 経路

#075：ラクダは砂漠を歩く。

#076：敵はこの道を通して逃げるはずだ。

行為の経路を示す項をとる自動詞文。「経路」がどのような格標識／側置詞で表されるか。

1.24 起点

#077：娘が樹の上から落ちてきた。

行為の起点を示す項をとる自動詞文。「起点」がどのような格標識／側置詞で表されるか。

#078：明日、われわれはこの基地から目的地まで飛ぶ。

行為の起点と着点の双方が現れる自動詞文。それぞれがどのような格標識／側置詞で表されるか。

1.25 原因

#079：雪が降ったので軍隊は出発を遅らせた。

原因となる出来事を示す表現をもつ文。外界にある事物が原因の場合。どのような標識で「原因」を示すか。

#080：観衆は大喜びをして堀の中に飛び込んだ。

原因となる出来事を示す表現をもつ文。感情が原因の場合。どのような標識で「原因」を示すか。

1.26 経験者主語

#081：この国の人々はみなビールが好きだ。

心理状態を表す二項述語を使った文。格表示のパターン／側置詞の使用パターンが典型的な他動詞の場合と同じか、異なるか。

#082：サルに芸術がわかるはずがない。

思考活動を表す二項述語を使った文。格表示のパターン／側置詞の使用パターンが典型的な他動詞の場合と同じか、異なるか。

1.27 時間の経過

#083：われわれは結婚して20年になる。

#084：釈迦が死んで2500年になる。

できごとの発生時点から現在までの経過時間を表す文。動詞のアスペクト形式でそれを表現するのか、専用の構文パターンを用いるのか。

1.28 比較

#085：紅葉した葉はバラの花よりも赤い。

2者を比較する文。「～よりも～」型の比較。比較の対象を導入する句は、述語に対してどの位置にあるか。

#086：北極は南極ほど寒くはない。

2者を比較する文。「～ほど～でない」型。比較の対象を導入する句は述語や否定辞に対してどの位置にあるか。

1.29 比喩

#087: 散る花びらが雪のように見える。

比喩を導入する文。比喩の部分が名詞句。どのような標識を用いるか。

#088: わたしはトリが飛ぶように海の中を泳いだ。

比喩を導入する文。比喩の部分が節になっている。どのような標識を用いるか。

1.30 引用

#089: 教師は「ともだちと仲良くしなさい」と言う。

直接話法の引用を含む文。どのような引用の標識を用いるか。

#090: ある人々は彼を正直者だと言う。

間接話法の引用を含む文。目的語の繰り上げ(例:「彼が」→「彼を」)が行われているか、いないか。

1.31 信念

#091: 熱心な信者は地球が丸いと信じている。

信念を表明する文。いわゆる「世界創造述語」を使った文。

#092: 同級生たちは彼が女性だと思っていた(がそうではなかった)。

事実とは異なる想像を述べる文。「事実ではない」を示す何らかの標識があるかどうか。

1.32 知識

#093: われわれは彼が宇宙人だということを知っていた。

知識を表現する動詞が節目的語をもつ文。

#094: あなたは彼がまだ生きていますかどうかわかりますか。

知識を表現する動詞が節目的語をもつ間接疑問文。どのような間接疑問文の標識があるか。

1.33 目的

#095: 父は月を見るために外に出た。

行為の目的を導入する節がどのような構造をもつか。完了したできごとの場合。「外に出て月を見た」という構造と同じか、異なるか。

#096：われわれは病気を治すために薬をのむ。

行為の目的を導入する節がどのような構造をもつか。未完了のできごとの場合。完了の場合と何かが異なるか。

#097：看護師は外が見えるようにカーテンを開けた。

行為の結果がどのような状態であることを望むかを述べる文。

1.34 継起的行為

#098：姉はテレビを一台買って、祖母に送った。

#099：祖父は外に出てタバコを吸った。

時間的に連続した行為を述べる文。行為2は行為1の目的とも解釈できる文。先後関係の標識が現れるかどうか。

1.35 名詞句の構造

#100：わたしのある(1人の)友人

「ある／1人の」という不定の標識と「わたしの」という限定語がどう併存するか。その位置関係を見る。

#101：わたしのこの友人

「この」という指示詞と「わたしの」という限定語がどう併存するか。その位置関係を見る。

#102：この狭い庭

指示詞と修飾語の位置関係がどのようになっているか。

#103：この狭すぎる庭

指示詞と修飾語の位置関係を見る。修飾語に強調語がついている場合、何かが異なるか。

#104：この庭が狭い家

関係節と中心語と指示詞の位置関係。

1.36 関係節

#105：花を摘む人

中心語が関係節の「動作者」項の場合。

#106：鳥を捕る網

中心語が関係節の「道具」項の場合。

#107：虫を飼う場所

中心語が関係節の「場所」項の場合。

#108：魚を焼く煙

中心語が関係節内に対応項をもたない構造。このような関係節を許容するかどうか。

#109：彼にはパリに留学している息子が2人いる。

制限的關係節。「息子」はこの2人以外にもいる可能性がある。

#110：彼には息子が2人いて、みなパリに留学している

「息子」はこの2人しかいない場合。「非制限的關係節」で表すのか、節の連続で表すのか。

2. 結

この100文の選定には、もちろん様々な言語学的研究の著作を参考にした。ここにいちいちその題名を挙げることはできないが、先行研究の成果に多くを負っていることは明記しておきたい。

なお、この「100の文」を実際に多言語に翻訳する作業は、現在、その試行段階として、100のうち10を選んで、幾つかの言語（標準華語、ロシア語、タイ語など）に翻訳し、問題点を洗い出す作業を進めているところである。もちろん、当面の目標は、100文すべてをより多くの言語に翻訳することであるが、このコーパスを単なる対訳資料集で終わらせないために、最終的には、個々の言語データに言語学情報をメタ情報として付与するマークアップを施す予定である（このマークアップに関しては、山崎2006参照）。このマークアップの方式の検討も並行して進行中である。

謝辞

この資料の作成に当たっては、科学研究費補助金（基盤研究（B））、課題番号：19300047、『LCTLを含む多言語平行マルチメディア資源の構築と構造化方式の研究』、代表者：堀 一成）の補助を受けている。

参考文献

- 山崎直樹. 2006. 「構造化デジタルテキストと言語類型論——メタ情報をもつコーパスの拓く可能性——」, in Harada, Y. & J-B. Kim (eds.) *Studies in Korean and Japanese Linguistics and Natural Language Processing*, WIT, Inc., Tokyo.