

初級フランス語学習者の読解における 語彙知識の予測

杉 山 香 織

はじめに

本研究は、フランス語の学習歴1年半から2年半の学習者を対象に、読解における語彙知識を予測するモデルを作成し、その妥当性を検証することを目的とする。語彙知識は複雑で多面的であるが、本研究では語彙知識の広さに焦点を当て、語彙頻度情報に基づくモデルを作成する。さらにその語彙頻度情報がモデルの作成に有効かを確認し、モデルが学習者の受容語彙レベルを予測できるかを検証する。

1. 読解と語彙知識

第二言語の読解には、多くの変数が直接的・間接的に関連し、読み手に複雑な言語能力を要する。その中でも、理解にとって最も基本的で重要なプロセスは、素早く自動的に単語を認識し、意味へアクセスすることである (Grabe and Stoller 2002)。このような単語が文章の中に多いほど、理解も高まる (Laufer 1989a, Hu and Nation 2000)。適切な読解理解を保証するために、どれだけの既知語が含まれるのかを検証した研究によると、55%の読解理解には95%の語彙カバー率を要し (Laufer 1989b)、85.7%の読解理解には98%の語彙カバー率が必要である (Hu and Nation 2000) とされている。

語彙知識自体も一元的ではない。Daller et al. (2007) は、語彙知識が学習者によってどのように獲得され、蓄積されるのかについて、広さ、深さ、流暢さの三次元で説明している。それぞれ、語彙知識の広さはどれくらいの数の単語

を知っているか、語彙知識の深さは単語についてどれくらい多くの知識があるのか、語彙知識の流暢さはどれくらい自動的に単語知識を引き出せるかを指している。

語彙知識を測定するために、これまで主に三種類の方法がとられてきた(Koda 2005)。一つ目は、多肢選択式によって意味を一致させるテストや翻訳テストなどの伝統的なテスト形式を用いた手法である。二つ目は、語彙の広さを測定する語彙サイズテストである。語彙リストの頻度情報に基づいて、どのくらいの頻度層の語彙知識があるのかを測定する。代表的なものとして、Nation (1990) や Schmitt, Schmitt, & Clapham (2001) による Vocabulary Levels Test や Meara and Milton (2003) の X-Lex がある。三つ目は、語彙の深さを測定するテストであり、Vocabulary Knowledge Scale (Wesche and Paribakht 1996) が広く使用されている。これは、「単語を見たことがない」という語彙知識レベルから、「その単語を文中で使用することができる」という産出レベルまでの五段階を設定し、そのレベルに応じて学習者の語彙知識を測定するものである。

この中でも、近年の外国語学習者における語彙知識に関する研究の多くは、語彙の広さに焦点を当てた語彙サイズテストに基づくものである。学習者の既知語の総量を測定することができる語彙サイズテストには、単語の頻度情報を付した語彙リストが必要である。先の Vocabulary Levels Test や X-Lex は英語の語彙頻度リストに基づくものである。

フランス語の語彙頻度情報を提供する代表的なものとして、*VocabProfil* と *FLELex* が挙げられる。*VocabProfil* は、*Frequency Dictionary of French* (Lonsdale et Le Bras 2009) と、2300万語からなる話し言葉と書き言葉コーパスに基づき、頻度情報を計算している。語の計算単位は単語家族である。頻度上位1000語ごとに一つの頻度層を形成し、25の頻度層を設定している。たとえば、K1という頻度層は単語家族の頻度上位1位から1000位までを含み、K2には1001位から2000位までの単語家族を含んでいる。*FLELex* はCEFR (ヨーロッパ言語共通参照枠: Common European Framework of Reference for Languages) に準拠した教科書や学習者向けの本のコーパスに基づいて頻度情報を算出している。そのため、単語の難易度をCEFRのレベルで示すことができる (François et al. 2014)。

2. 研究設問

本研究の目的は、*VocabProfil* と *FLELex* から得られるフランス語の語彙頻度情報を用いて初級学習者の受容語彙知識を予測することである。そこで以下の2点について分析を行っていく。

- Q1. 語彙頻度情報をもとに作成したモデルは、学習者の既知語と未知語をどれくらい正しく判別できるか。
- Q2. 正しく判別できた場合、それらの単語にはどのような特徴があるか

3. 研究方法

3.1 調査参加者

本研究では、フランス語を主専攻とする大学2年生から3年生の10名を対象とする。調査時点では、フランス語学習歴はそれぞれ1年半から2年半であった。学習歴および取得資格は以下のとおりである（表1）。なお、いずれの学生も調査時点では留学経験を持たないが、調査時点以降に留学することが決定している。

表1：調査参加者の学習歴と資格

ID	学習歴 (年)	資格
1	1.5	DELFA1
2	1.5	DELFA1
3	2.5	DELFA2
4	2.5	DELFA2
5	2.5	DELFA2
6	2.5	DELFA2
7	2.5	DELFA2
8	2.5	DELFA2
9	2.5	DELFA2
10	2.5	DELFA2

3.2 調査する単語リストの選定

まず、調査するテキストとして、DELFL 対策の問題集の中から実際の試験形式を踏襲した A1 レベルから B1 レベルの模擬試験 1 回分を選択した。実際の形式と同様、模擬試験は A1 と A2 レベルについてはそれぞれ 4 テキスト、B1 レベルについては 2 テキストから構成されている。杉山 (2018) の調査に基づき、学習者が一人でも未知語であると判断した単語をすべてリスト化した。ただし、成句や熟語表現は頻度情報を得られないため、リストから除外する。また、杉山 (2019) の結果から、動詞の活用形は動詞自体の語彙の頻度が高くても習得されにくいことが明らかになったため、リストから除外することとした。同様に、代名動詞も他動詞からの類推に失敗し、既知の形態素を組み合わせで誤った意味推測を導く語 (deceptive transparent words) となるケースが多く見られたため、これらもリストから外した。各レベルのテキストにおける異なり語数と、リストに含まれる語数は以下のとおりである。

表 2: テキストにおける異なり語数とリスト内語数

レベル	異なり語数	リスト内語数
A1	209	62
A2	273	75
B1	441	121

3.3 語彙知識の測定とグループ化

リストに含まれた 258 語について、調査参加者はテキストを読みながら意味を記入した。各単語について意味が正しいかどうかを筆者が確認し、正しい場合は既知語、正しくない場合や空欄の場合は未知語と判定した。

その後、調査参加者をケース、判定を変数に用いたクラスター分析¹を行い、既知語に分類される傾向にある単語と未知語に分類される傾向にある単語に分けた。その結果、60 語が既知語、198 語が未知語と分類された。

¹ ユークリッド距離、ワード法

3.4 モデルの作成

本調査では、既知語と未知語を予測するモデルの作成に語彙頻度情報を使用する。今回使用するのは、*VocabProfil*と*FLELex*の2つである。*VocabProfil*は上記で述べたとおり2500万語規模の書き言葉と話し言葉のコーパスから頻度情報を得ており、フランス語圏で生活する上で触れる語彙の出現頻度を反映している。一方、*FLELex*は教科書コーパスから頻度情報を得ているため、学習環境で触れる語彙の出現頻度を反映している。

*VocabProfil*は各単語が属する頻度層についての情報を提供するが、杉山(2018, 2019)から頻度2001位から未知語の割合が大幅に増えるということが明らかになっているため、頻度1000位まで、1001位から2000位まで、それ以降の3グループに分けた。

*FLELex*は各単語の正規化された頻度情報を提供するため、*VocabProfil*の頻度情報と性質が異なる。したがって、*VocabProfil*に合わせて正規化された頻度の高い順に1000位まで、1001位から2000位まで、それ以降の3グループに分けた。*VocabProfil*と*FLELex*の頻度情報を掛け合わせたクロス表は以下のとおりである。

表3: *VocabProfil*と*FLELex*のクロス表 (語数)²

	FL1	FL2	FL3
VP1	50	18	18
VP2	10	19	17
VP3	6	19	101

表3の線で囲まれた50の単語は、*VocabProfil*と*FLELex*ともに頻度が一番高い層に分類された単語であるため、一般的な使用頻度も高く、学習環境で触れる頻度も高い語だといえる。一方、破線で囲まれた101の単語は*VocabProfil*と*FLELex*ともに頻度が一番低い層に分類された単語であるため、一般的な使

² VPは*VocabProfil*、FLは*FLELex*、1は1000位までの頻度層、2は1001位から2000位までの頻度層、3はそれ以降を指す。

用頻度も学習環境における頻度も低い単語である。直感的に、前者の方が習得されやすい単語であるといえる。表4はそれぞれ、高頻度層（VP1かつFL1）と低頻度層（VP3かつFL3）に分類される単語の例である。

表4：高頻度層と低頻度層に分類される単語の例

VP1∧FL 1	VP3∧FL 3
accepter (受け入れる)	abbaye (修道院)
activité (活動)	abordable (手頃な)
apporter (持っていく)	affirmation (断言)
association (団体)	alentour (周囲に)
aucun (何も...ない)	alerte (警告)
avenir (未来)	autoroute (高速道路)

VocabProfil と *FLELex* の頻度情報を説明変数、3.3でクラスター分析によって分類された「既知語」と「未知語」を目的変数として線形判別分析を行い、Fisher の線形判別関数をもとめる。

4. 結果と考察

4.1 線形判別分析結果

未知語と既知語を予測する Fisher の線形判別関数は、それぞれ以下の通りとなった。

$$\begin{aligned} \text{未知語} &= 1.387 \times \text{VP 頻度層} + 1.890 \times \text{FL 頻度層} - 3.407 \\ \text{既知語} &= 1.373 \times \text{VP 頻度層} + 3.414 \times \text{FL 頻度層} - 6.484 \end{aligned}$$

この関数にしたがって分類された結果は表5の通りである。全体の判定率は68.20%であった。既知語についてはあまり正しく判定することができなかった。関数によって既知語と判定された122語のうち、クラスター分析でも既知語と分類されたものは50語にとどまった。

一方、未知語はかなりの割合で正しく予測できた。関数で未知語と判定された136語のうち、クラスター分析でも未知語に分類されていたものは126語であった(表5)。

表5：分類結果

		予測		
		既知語	未知語	合計
クラスター	既知語	50	10	60
	未知語	72	126	198
	合計	122	136	258

判定関数に基づいて未知語の判定パターンをまとめると、以下の通りとなる。

表6：未知語の判定パターン

	VP	FL
①	1	3
②	2	3
③	3	3

未知語の判定パターンは① VP1かつ FL3② VP2かつ FL3③ VP3かつ FL3となり、*FLELex*の頻度の低さが学習者の未知語の要因となっていることが分かる。すなわち、学習環境で触れる頻度の低い単語が未知語である確率が高いということになる。

4.2 判定エラー

未知語の判定率が高かったものの、いくつか判定エラーがあった。上記の① (VP1かつ FL3) に分類された18単語については、*pratique* 以外はすべて未知語に判定されている。

moyenne, immobilier, note, réforme, règlement, auteurs, commune, établissement, déclaration, élément, pression, volonté, **pratique**, engagement, essai, venue (名詞), commune, tribunal

② (VP2かつ FL3) については、17語中 *institut* のみが判定エラーであった。

disponible, affiche, port, trente, interdit, critiquer, **institut**, initial, hors, épouse, uniquement, critère, démontrer, bilan, étape, protocole, constitution

③ (VP3かつ FL3) は101語が該当するが、そのうち *célébration*, *invitation*, *officiellement*, *orange*, *participants*, *studio*, *symbolique*, *talent* の8語が判定エラーであった³。

abbaye, abordable, affirmation, alentour, alerte, autoroute, bague, baigneur, balade, balcon, ballon, balnéaire, basilique, bassin, bougie, boulot, cases, catimini, **célébration**, centenaires, chandelle, compteur, concocté, conducteur, conjointe, conjointement, conjugale, conjugalité, consolider, courriel, crevette, croix, décès, déchetterie, démographique, dicton, duplex, emménager, enceinte, expérimental, exploration, falaise, ferré, fiançailles, formellement, gastronomique, gratuitement, greffe, grilles, guichet, hétérosexuel, horizon, hôtellerie, huître, idéal, idéal, imposition, indispensable, initialement, insérer, instance, intime, **invitation**, itinéraire, jury, location, matrimoniale, **officiellement**, orange, **orange**, pacs, pacsée (名詞), panneaux, **participants**, partielle, pèlerin, perception, possession, poubelle, probatoire, progressive, prudent, rail, rarement, record, résidence, restaurer, rubrique, sensation, signification, stand, strictement, **studio**, **symbolique**, **talent**, thèse, tradi, trajectoire, trajet, vaccin, vestiaires

判定エラーとなったこれらの単語のうち *pratique* については、①で分類されている単語の中でもジャンルや話題に囚われない汎用性の高い単語であるため、既知語となった可能性がある。*studio* については、全ての調査参加者が留学予定者であり、フランス語で書いてある住居関係の資料に触れたことがあるため、その意味を知っていたと推察できる。その他の単語については、英語に似た形の単語があり、意味も英語とあまり変わらないため、英語の語彙知識として持っている可能性が高いものである (表7)。

³ 同じ単語が複数含まれるものもあるが、これは異なるテキストや異なる場所に当該単語が使われていたケースである。

表7：判定エラーの単語

フランス語	英語
institut	institute
célébration	celebration
invitation	invitation
officiellement	officially
orange	orange
participants	participants
symbolique	symbolic
talent	talent

4.3 未知語の特徴

前節では、*FLELex*において頻度の低い単語であっても、英語の語彙知識があることによってフランス語の意味を導き出すことのできた例を見てきた。一方で、未知語のリストの中には、英語と全く同じ形や極めて類似した形を持つフランス語の単語も多く存在した。表8はそのような単語の一例である。

表8：フランス語と英語の対照⁴

フランス語	英語	意味	フランス語	英語	意味
réforme	reform	改革	formellement	formally	正式に
commune	commune	共同社会	gastronomique	gastronomic	美食の
déclaration	declaration	宣言	grilles	grilles	グリル
élément	element	要素	hétérosexuel	heterosexual	異性愛の
engagement	engagement	婚約	horizon	horizon	地平線
tribunal	tribunal	裁判所	idéel	ideal	理想
port	port	港	imposition	imposition	税金
initial	initial	始めの	indispensable	indispensable	不可欠な
critère	criteria	基準	initialement	initially	始めに
abbaye	abbey	大修道院	instance	instance	訴訟
affirmation	affirmation	断言	itinéraire	itinerary	旅程
alerte	alert	警報	jury	jury	審査員

⁴ ここに示した意味は、本調査で使用したテキストの文脈における意味である。

フランス語	英語	意味	フランス語	英語	意味
autoroute	autoroute	高速道路	matrimoniale	matrimonial	婚姻の
balcon	balcony	バルコニー	partielle	partial	一部の
basilique	basilica	大聖堂	perception	perception	知覚
bassin	basin	流域	possession	possession	所有
bougie	bougie	ろうそく	prudent	prudent	慎重な
centenaires	centenary	100年の	rail	rail	線路
conducteur	conductor	車掌	résidence	residence	住居
conjugale	conjugal	結婚の	restaurer	restore	再建する
conjugalité	conjuality	婚姻	sensation	sensation	感覚
consolider	consolidate	強固にする	signification	signification	意味
démographique	demographique	人口統計の	stand	stand	売り場
duplex	duplex	メゾネット	strictement	strictly	厳しく
enceinte	enceinte	妊娠している	thèse	thesis	論文
expérimental	experimental	実験の	trajectoire	trajectory	道筋
exploration	exploration	調査	vaccin	vaccine	ワクチン

これらの単語は、そもそも英語の語彙知識として持ち合わせていない可能性が高い。したがって、フランス語の語彙知識は英語の語彙知識にも大きく左右されることが示唆できる。

また、学習者の英語の語彙知識や和製英語から類推を試みたものの、本テキストの文脈では誤った意味となってしまった例や、正しい意味を導くことのできなかった例も多く見られた。

表9：英語からの類推に失敗した例

フランス語	意味	英語	意味
cases	枠	cases	ケース
conjoint	配偶者	conjoint	結合した
initial	はじめの	initial	イニシャル
location	賃貸、リース	location	ロケーション、場所
note	メモ	note	ノート
progressive	段階的な	progressive	進歩的な
réforme	改革	reform	*リフォーム
uniquement	単に	uniquely	比類なく

*リフォームは和製英語

5. まとめと今後の課題

以上から、リサーチクエスションの答えは以下の通りとなる。

Q1. 語彙の頻度情報をもとに作成したモデルは、学習者の既知語と未知語をどれくらい正しく判別できるか

Fisher の線形判別分析を用いて学習者の既知語と未知語の予測モデルを検証したところ、判定率は68.20%であった。既知語と判定されたものの中には学習者にとっての未知語が多く含まれていたが、未知語についてはかなり正しく予測することができた。

Q2. 正しく判別できた場合、それらの単語にはどのような特徴があるか

正しく判別できたのは未知語であるが、未知語に分類される単語は教科書コーパスにおいて頻度の低い単語であった。日本国内のみでフランス語を学習する学習者にとって、教科書はインプットの中心であり、学習環境で触れる単語が低ければ習得される可能性が低いということを再確認することができた。判別関数では未知語と判定されたが、学習者の既知語であった判定のエラーの単語については、英語に類似した形があり、その英語の意味を習得済みであるものが多く含まれていた。一方で、英語と同形か、極めて類似した形を持つ単語でも、英語の語彙知識の欠如や、英語の意味からの類推に失敗したことにより、未知語に分類されたものも多い。

今後の課題としては、以下の点が挙げられる。まず、判定関数によって既知語と分類された単語の多くは、実際には学習者にとっての未知語であった。そのため、既知語の判定エラーとなった単語についても分析をしていかなければならない。

次に、留学経験を持たない初級学習者の未知語をある程度高い精度で予測できたが、調査参加者数も初級学習者に限定されており10名と少ない。そのため、今回対象とした学習者とは別の初級学習者にもこの結果が適応できるのかを追調査する必要がある。また、留学経験を持つ中上級の学習者にもこの判定モデルが適応できるのかについて調査しなければならない。

また、本研究で使用した変数は *VocabProfil* と *FLELex* の頻度情報のみであったため、変数をもっと増やし、判定モデルを精緻化しなければならない。

たとえば、未知語に分類された単語の中には、英語の語彙知識と深く関連するものがあつた。そのため、英語知識も変数に入れるべきかどうかの検討を慎重にする必要がある。

さらに、今回使用したのはA1レベルからB1レベルにおける1回分のリーディング試験のテキストのみであるため、調査した単語の数が限られている。本調査で用いたテキストとは別のテキストについても、本判定式が有効であるのかを検証しなければならない。

最後に、本研究結果を教育現場へ還元する実践方法を考える必要がある。留学経験のないフランス語学習者にとって、教科書など学習環境下で触れる単語が語彙能力に大きな影響を与えることが再確認ができた。そのため、初級学習者には *FLELex* の高頻度語を含むテキストをまず選択し、インプットの機会を多く与えることで、高頻度語を優先的に習得させることが求められる。一方で、テキストレベル自体は高くないものであっても、低頻度語が多く含まれていた。本調査で使用した単語のうち、A1レベルのテキストには19語、A2レベルのテキストには41語、B1レベルのテキストには76語が、*FLELex* のコーパスにおいて低頻度語に分類されたものであつた。A1レベルの異なり語数は209であるため、その中の9.09%が低頻度語であり、A2レベルについては異なり語数が273であるのに対して15.02%が、B1レベルについては異なり語数441に対して17.23%が低頻度語に相当する。*FLELex* の頻度情報は教科書コーパスから得られているものの、コーパスに含まれるテキストのジャンルや話題によっては頻度が大きく左右される単語もある。このような場合、ジャンルに関する背景知識や英語の知識が、フランス語の単語の意味判別に大きく役立つことがある。そのため、既に持っているこれらの知識を素早く引き出し、テキストの中でフランス語の語彙知識として関連付け、類推できるように手伝えるタスクを考えていく必要がある。

参考文献

- DALLER, H., MILTON, J. & TREFFERS-DALLER, J.(2007). *Modelling and assessing vocabulary knowledge*. Cambridge : Cambridge University Press.
- FRANÇOIS, T., GALA, N., WATRIN, P. & FAIRON, C.(2014). FLELex: a graded lexical resource for French foreign learners. In *the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. Reykjavik, Iceland, 26-31 May.
- GRABE, W. & STOLLER, F.(2002). *Teaching and researching reading*. New York : Longman.
- HU, H. & NATION, I.S.P.(2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language* 13. pp. 403-430.
- LAUFER, B.(1989a). A factor of difficulty in vocabulary learning: Deceptive transparency. *AILA Review* 6. pp. 10-20.
- LAUFER, B.(1989b). What percentage of text lexis is essential for comprehension. In C. Lauren and M. Nordman(Eds.), *Special language: From humans thinking to thinking machines*. Clevedon: Multilingual Matters. pp. 316-323.
- LONSDALE, D. & LE BRAS, Y.(2009). *A frequency dictionary of French : core vocabulary for learners*. New York : Routledge.
- MEARA, P. M., & MILTON, J.(2003). *X_Lex. The swansea levels test*. Newbury : Express.
- NATION, I.S.P.(1990). *Taching and learning vocabulary*. Boston : Heinle an Heinle.
- KODA, K.(2005). *Insights into second language reading*. New York : Cambridge University Press.
- SCHMITT, N., SCHMITT, D. & CLAPHAM, C.(2001) Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing* 18. pp. 55-88.
- WESCHE, M. & PARIBAKHT, T.S.(1996). Assessing second language vocabulary knowledge: depth vs. breadth. *Canadian Modern Language Review* 53. pp. 13-39.
- 杉山香織(2018).「フランス語初中級学習者の受容語彙知識」,『外国語教育研究21』外国語教育学会紀要 . pp. 54-72.
- 杉山香織(2019).「フランス語学習者における受容能力の経年変化」,『外国語教育研究22』外国語教育学会紀要 . pp. 41-59.