

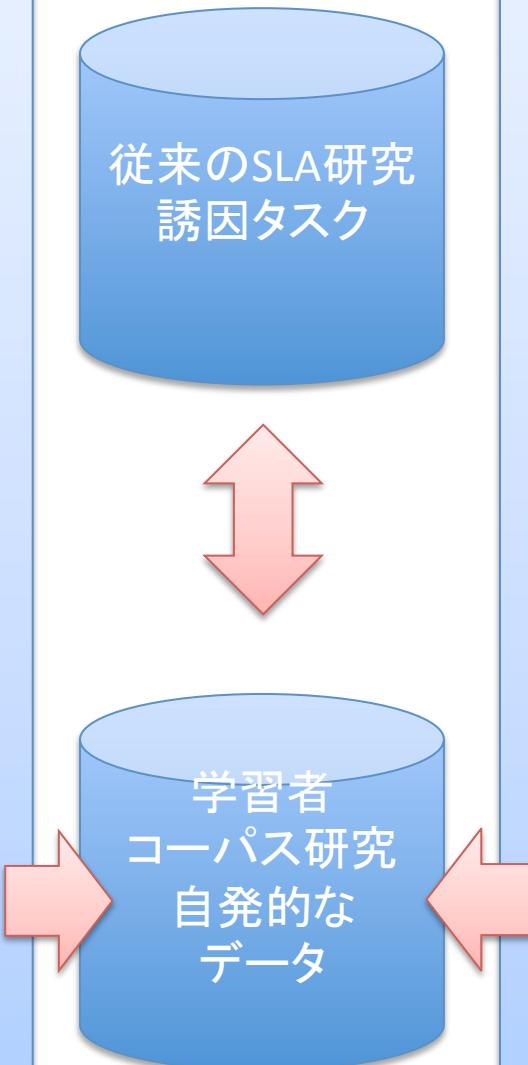
**コーパス分析に基づくCEFR基準特性としての
関係代名詞に関する研究：
学習者・コースブックコーパスを用いて**

東京外国語大学大学院
博士前期課程2年
高橋有加

研究の背景

ヨーロッパ言語共通 参照枠(CEFR)の 影響:

- 文科省:
CAN-DO リストの目標・評価への利用
- CEFR-Jなどの日本の英語教育への応用
- English Profile:
CEFR レベルの基準特性(criterial feature)の特定



関係節:

- 名詞の後置修飾(postnominal modification)の1種
- 日本人には習得の困難な文法事項
- SLA でもさまざまな習得順序仮説がある(Zobl 1989; Hamilton 1994; Izumi 2003; Shirai & Ozeki 2007)
- English Profile の research agenda にも関係節の基準特性としての有効性を検証する仮説があげられている(Hawkins 2009)

先行研究(1): 関係詞の分類

関係代名詞の機能による分類:

Subject (SU)	the professor [who wrote the letter] the book [which sold 1 million copies]
Direct object (DO)	the professor [whom the student knows] the book [which I bought yesterday]
Indirect object (IO) Oblique (object of preposition)	the professor [to whom the student showed the book] the professor [about whom we spoke] the book [which I told you about]
Genitive (GEN)	the professor [whose son the student knows] the book [whose title I forgot]
Object of the comparative (OCOMP)	the professor [that his wife is taller than] the book [that my edition is older than]

先行研究(2)：関係詞の分類

埋め込み位置と関係詞の種類による分類

SS	[_S [The dish] [[_S which] fell on the floor]] broke in half.
SO	[_S [The sweater] [[_O which] I found on the bus]] belongs to Susie.
OS	The little girl is looking for [_O [the cat] [[_S which] ran away]]
OO	The boys are reading [_O [the books [[_O which] they borrowed from the library]].

先行研究(3): 習得仮説

- Noun Phrase Accessibility Hierarchy (NPAH) Hypothesis
 - Comrie & Keenan (1979): typological universal の1例
 - SLA でもこの階層に準じた習得難易度があるという仮説 (Zobl 1989; Izumi 2003; Shirai and Ozeki 2007)
 - SU > DO > IO > OBL > GEN
- SO Hierarchy Hypothesis (Hamilton 1994)
 - 関係節の処理および習得の難易度は関係節の埋め込み位置および埋め込みの深さに影響を受ける
 - OS > OO/SS > SO
 - 埋め込み位置 : 目的語 > 主語
 - 関係代名詞のタイプ : 主格 > 目的格

研究設問

- Q1) 関係詞の頻度はレベルが上がるにつれて高くなるか。また、頻度は特定のCEFRレベルを特徴付ける基準特性になり得るか。
- Q2) 学習者のCEFRレベルごとの関係節の頻度はNPAHとSOHHに示された難易度に従うか。
- Q3) 表層系、NPAH, SOHHの分類において、学習者の使用分布とコースブックの出現頻度の傾向に類似点・相違点はあるか。
- Q4) 各レベルにどのようなエラーがみられるか。書き言葉と話し言葉の間に違いは見られるか。

使用したコーパス

Japanese EFL Learner (JEFLL) Corpus

- 20分間のin-class essay。辞書なし。
- 中学1年生～高校3年生
- 10,063のエッセイ
- 669,281語
- CEFRレベル情報付version

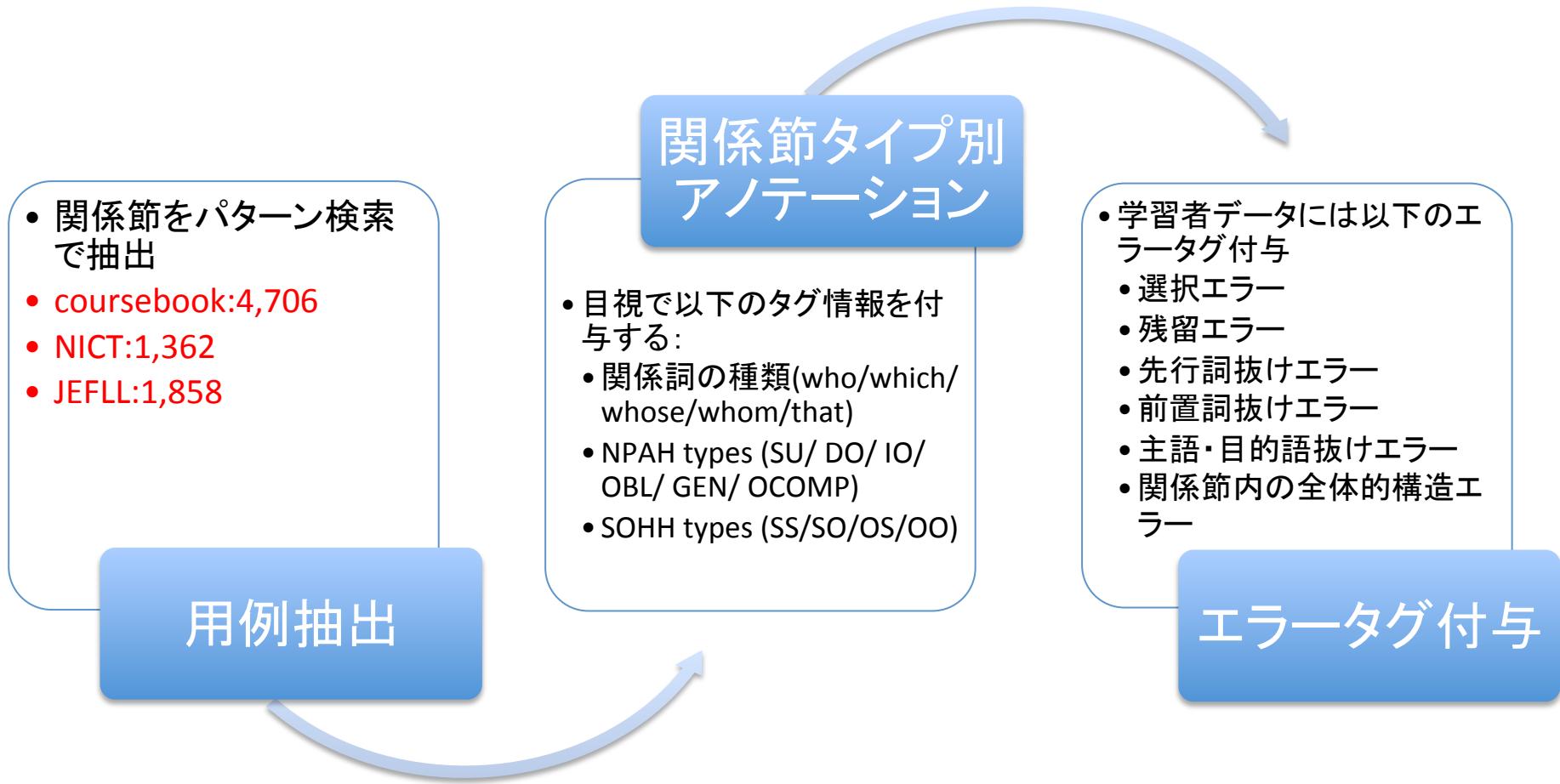
NICT JLE Corpus

- 15分間の口頭インタビューテストのスクリプト
- 1,281のサンプル；
2,176,000語
- 元のテストの9レベルをCEFR レベルに換算

CEFR Coursebook Corpus

- ヨーロッパで出版されているCEFR準拠のコースブック100冊
- 1,761,520語

データ処理の手順



関係節エラーの分類

- (SEL) Selection of relative pronouns 選択エラー

Our class had a drama *which* name is “Unexpected Guest”.

- (RSP) Resumptive relative pronouns 代名詞残留

I take out the thing which *it* is important to me.

- (MPR) Missing prepositions 前置詞欠落

He went to the place which he used to live ● .

- (MAT) Missing antecedents 先行詞欠落

At that time, ● who eat breakfast with me is my mother.

I can't believe ● who don't eat breakfast.

- (MRS) Missing subjects in relative clauses 関係詞節内主語欠落

I can meet my friends who ● haven't seen a long time.

- (MRO) Missing objects in relative clauses 関係詞節内目的語欠落

Our play was popular among the people who came and see ● .

- (STR) Ungrammatical structures in relative clauses 語順・構造のエラー

We sang a song which is a famous singer.

Q1

関係詞の頻度はレベルが上がるにつれて高くなるか。
また、頻度は特定のCEFRレベルを特徴付ける基準特性になり得るか。

結果

Percentage of relative clauses in all the instances of surface forms in JEFLL

Surface forms	Total occurrences	RP use	%
that	4271	420	10%
which	944	830	89%
who	557	485	87%
whose	46	46	100%
whom	17	17	100%

Percentage of relative clauses in all the instances of surface forms in NICT JLE

Surface forms	Total occurrences	RP use	%
that	6718	194	3%
which	960	509	53%
who	676	627	93%
whose	24	22	92%
whom	9	9	100%

一人当たりの使用率

The usage rate of overall RCs across CEFR levels in JEFLL

JEFLL	A1	A2	B1	B2
Number of files (people)	3597	4956	1529	46
RC sentences	129	861	794	23
Use rate per person	0.04	0.17	0.52	0.5

The usage rate of overall RCs across CEFR levels in NICT JLE

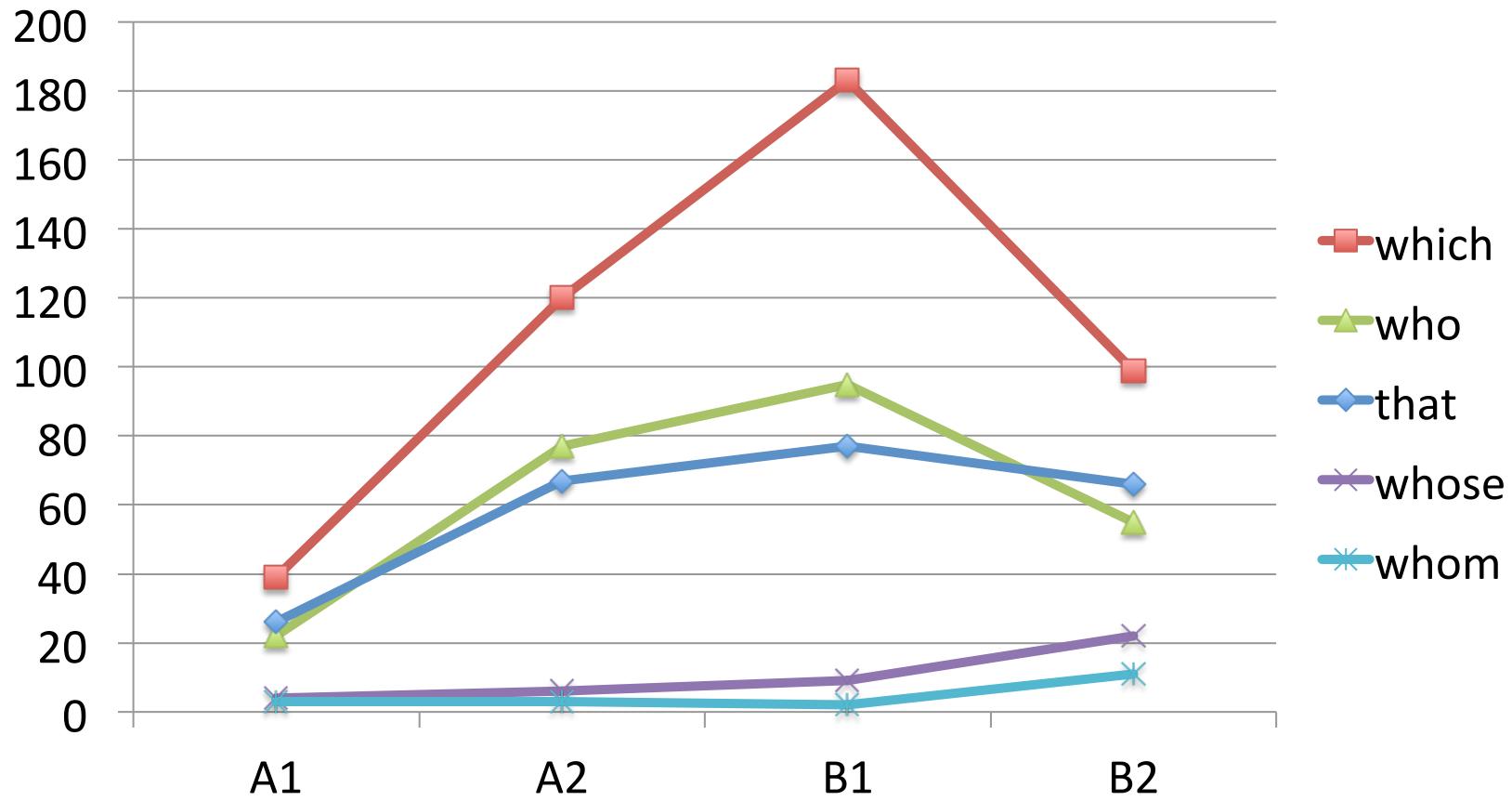
NICT	A1	A2	B1	B2
Number of files (people)	260	718	263	40
RC sentences	38	528	605	190
Use rate per person	0.15	0.74	2.3	4.75

JEFLL

Normalized frequencies (100,000 per words) of each surface form along the CEFR levels (JEFLL)

RP	A1	A2	B1	B2	Total
that	26	67	77	66	236
which	39	120	183	99	441
who	22	77	95	55	249
whose	4	6	9	22	41
whom	3	3	2	11	19
Total	94	272	366	254	986

JEFLL



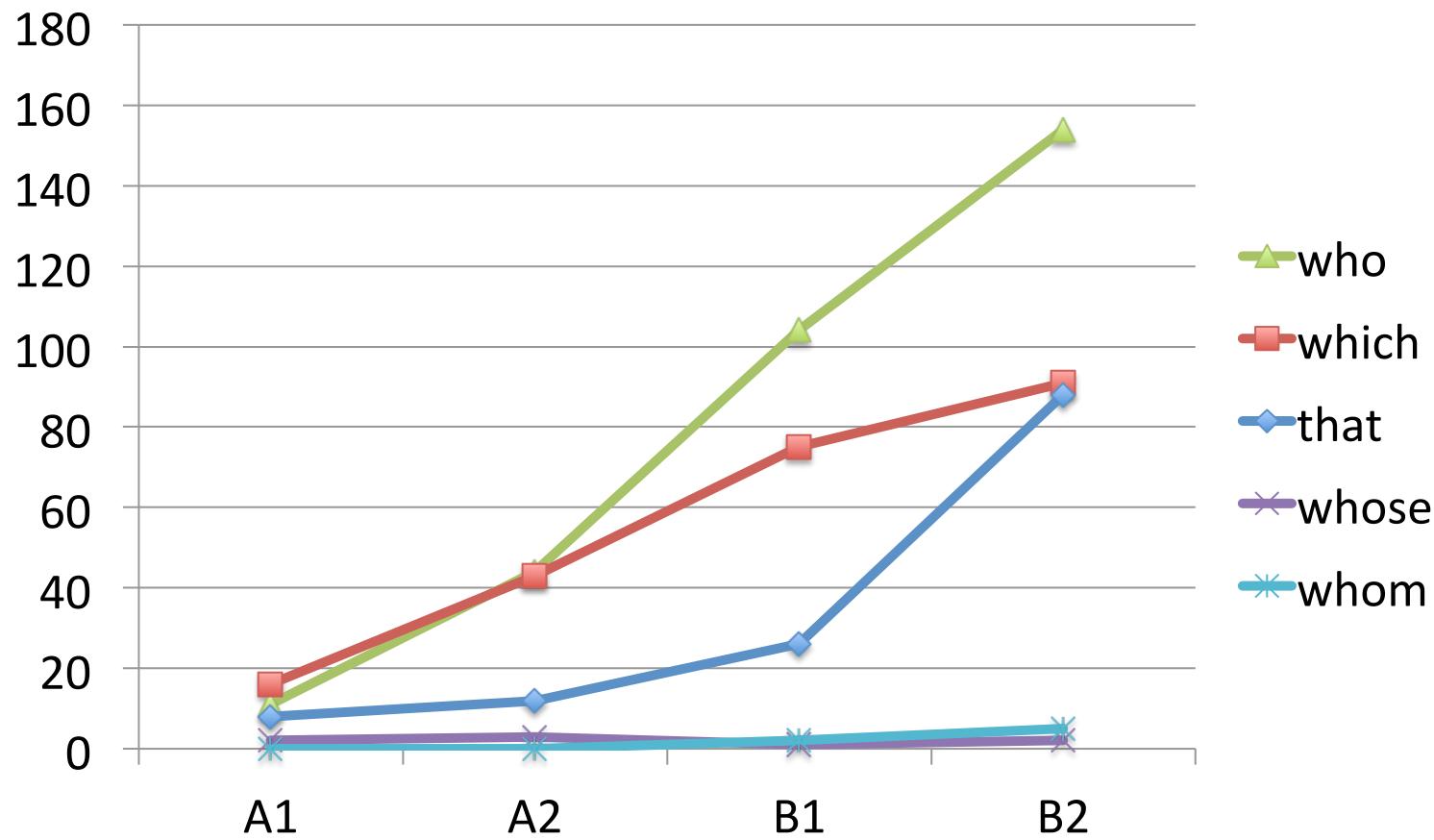
Normalized frequencies (100,000 per words) of each surface form along the CEFR level (JEFLL)

NICT

Normalized frequencies (100,000 per words) of each surface form along the CEFR levels (NICT JLE)

RP	A1	A2	B1	B2	Total
that	8	12	26	88	134
which	16	43	75	91	225
who	11	44	104	154	313
whose	2	3	1	2	8
whom	0	0	2	5	7
Total	37	102	208	340	687

NICT



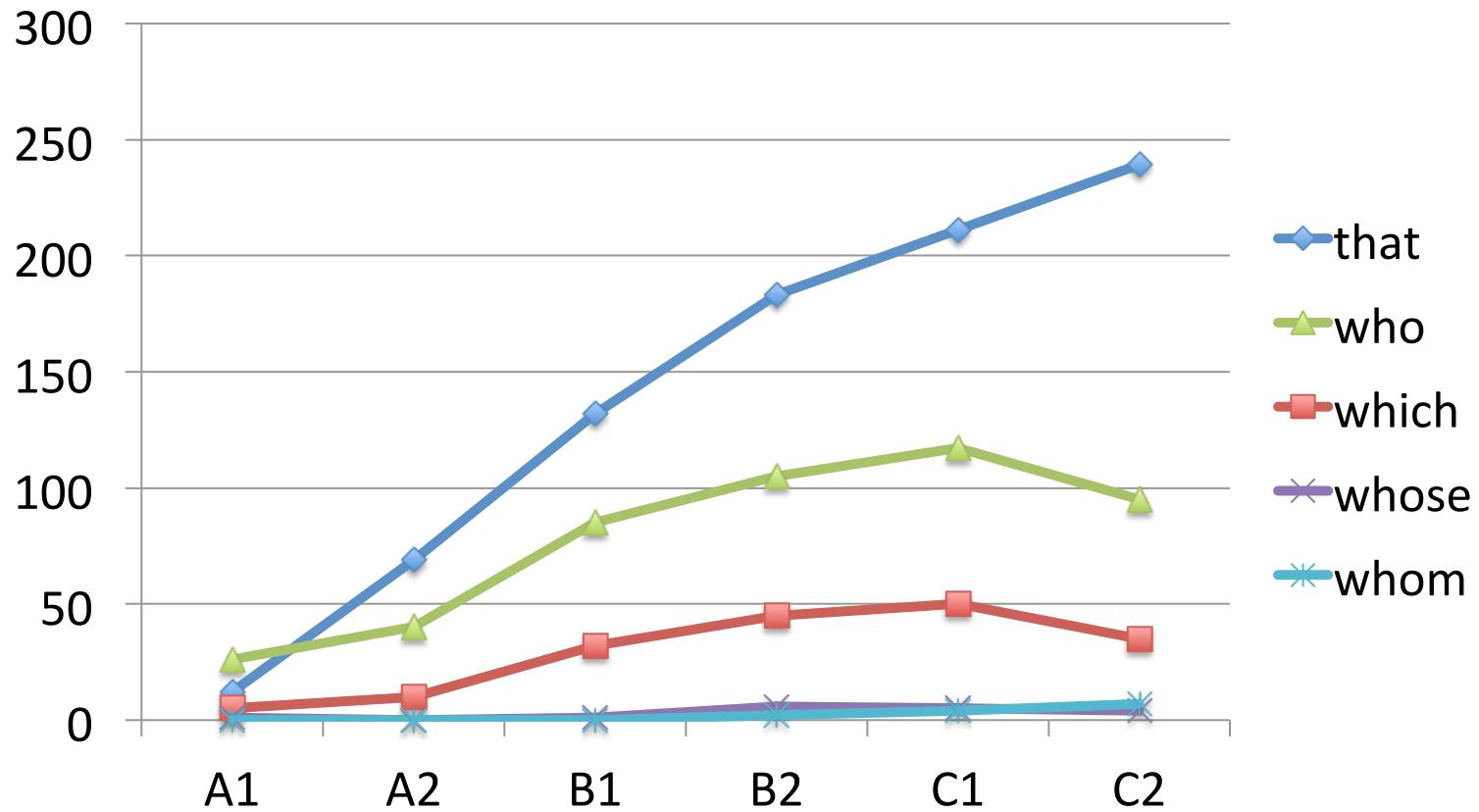
Normalized frequencies (100,000 per words) of each surface form along the CEFR levels (NICT JLE)

Coursebook

Normalized frequencies (100,000 per words) of each surface form along the CEFR levels (Coursebook)

CB	A1	A2	B1	B2	C1	C2	Total
that	12	69	132	183	211	239	846
which	5	10	32	45	50	35	177
who	26	40	85	105	117	95	468
whose	1	0	1	6	5	4	17
whom	0	0	0	2	4	7	13
Total	44	119	250	341	387	380	1521

Coursebook



Normalized *frequencies* (100,000 per words) in each surface form along the levels (Coursebook)

Q2

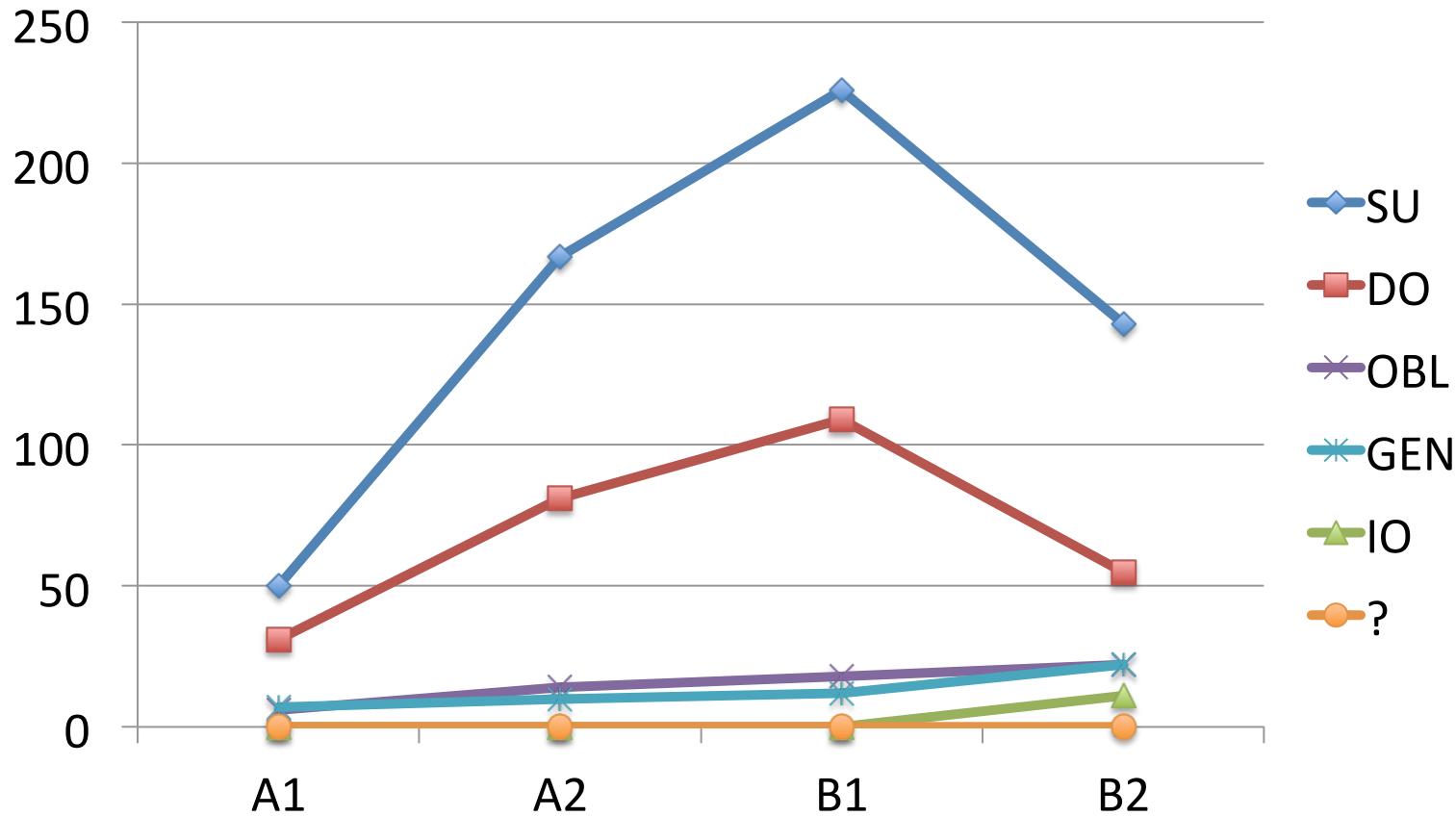
学習者のCEFRレベルごとの関係節の頻度はNPAHとSOHHに示された難易度に従うか。

NPAH (JEFLL)

Distributions of NPAH types across the CEFR levels (JEFLL)

NPAH	A1	A2	B1	B2	Total
SU	50	167	226	143	586
DO	31	81	109	55	276
IO	0	0	0	11	11
OBL	6	14	18	22	60
GEN	7	10	12	22	51
Unknown	0	0	0	0	0
Total	94	272	365	253	984

NPAH (JEFLL)



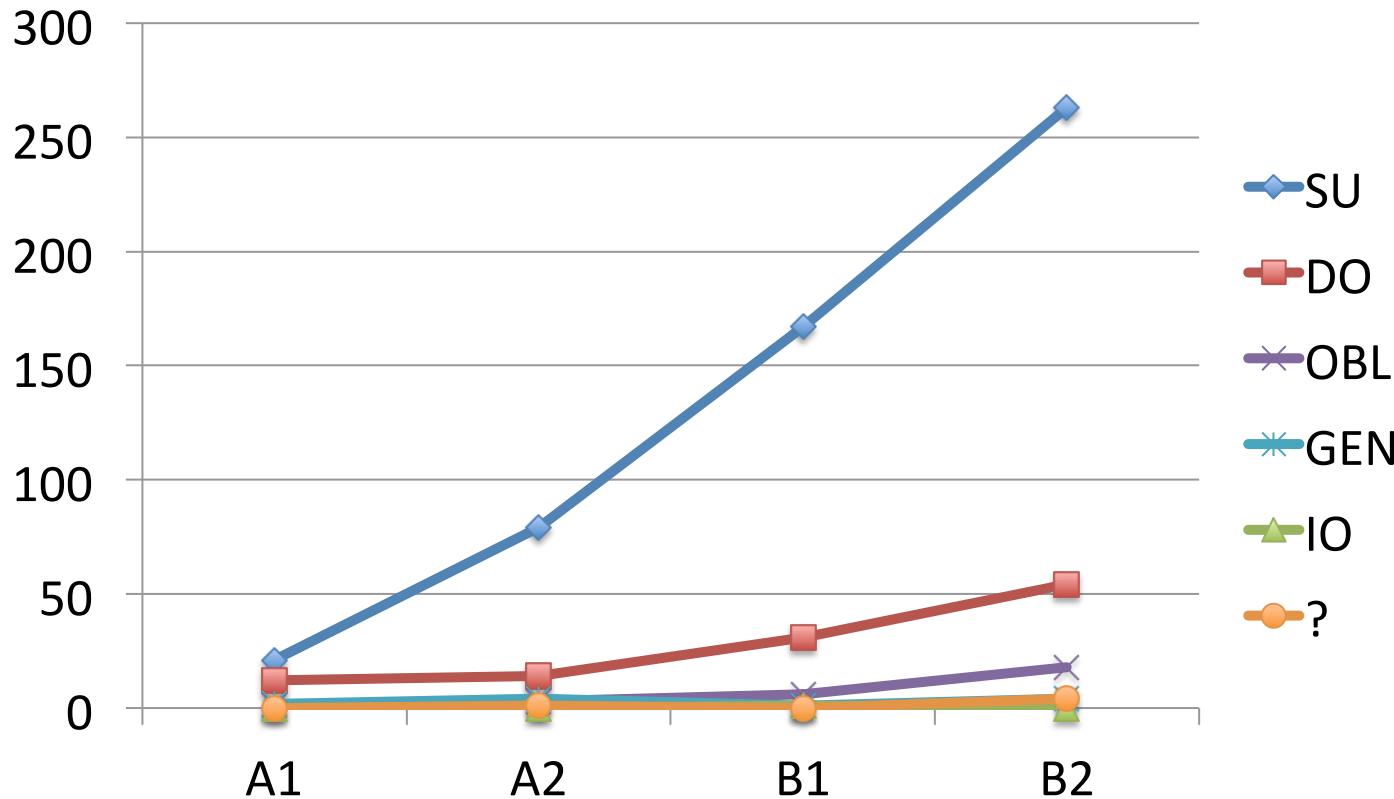
Distributions of NPAH types across the CEFR levels (JEFLL)

NPAH (NICT)

Distributions of NPAH types across the CEFR levels (NICT JLE)

NPAH	A1	A2	B1	B2	Total
SU	21	79	167	263	530
DO	12	14	31	54	111
IO	0	0	1	0	1
OBL	1	3	6	18	28
GEN	2	4	1	4	11
Unknown	0	1	0	4	5
Total	36	101	206	343	686

NPAH (NICT)



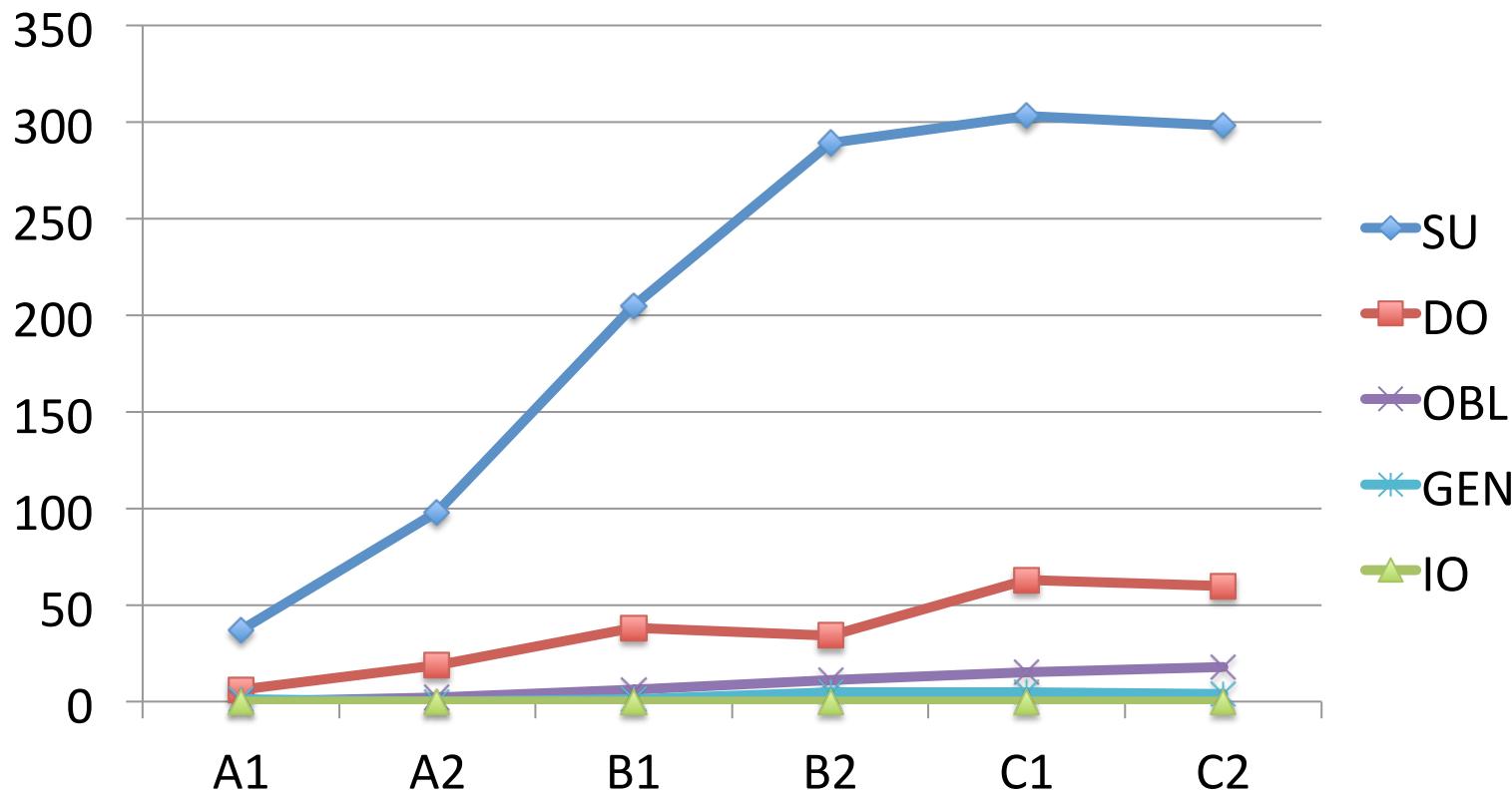
*Distributions of NPAH types across the CEFR levels
(NICT JLE)*

NPAH (Coursebook)

*Distributions of NPAH types across the CEFR levels
(Coursebook corpus)*

NPAH	A1	A2	B1	B2	C1	C2	Total
SU	37	98	205	289	303	298	1230
DO	6	19	38	34	63	60	220
IO	0	0	0	0	0	0	0
OBL	0	2	6	11	15	18	52
GEN	1	0	1	5	5	4	16
Total	44	119	250	339	386	380	1518

NPAH (Coursebook)



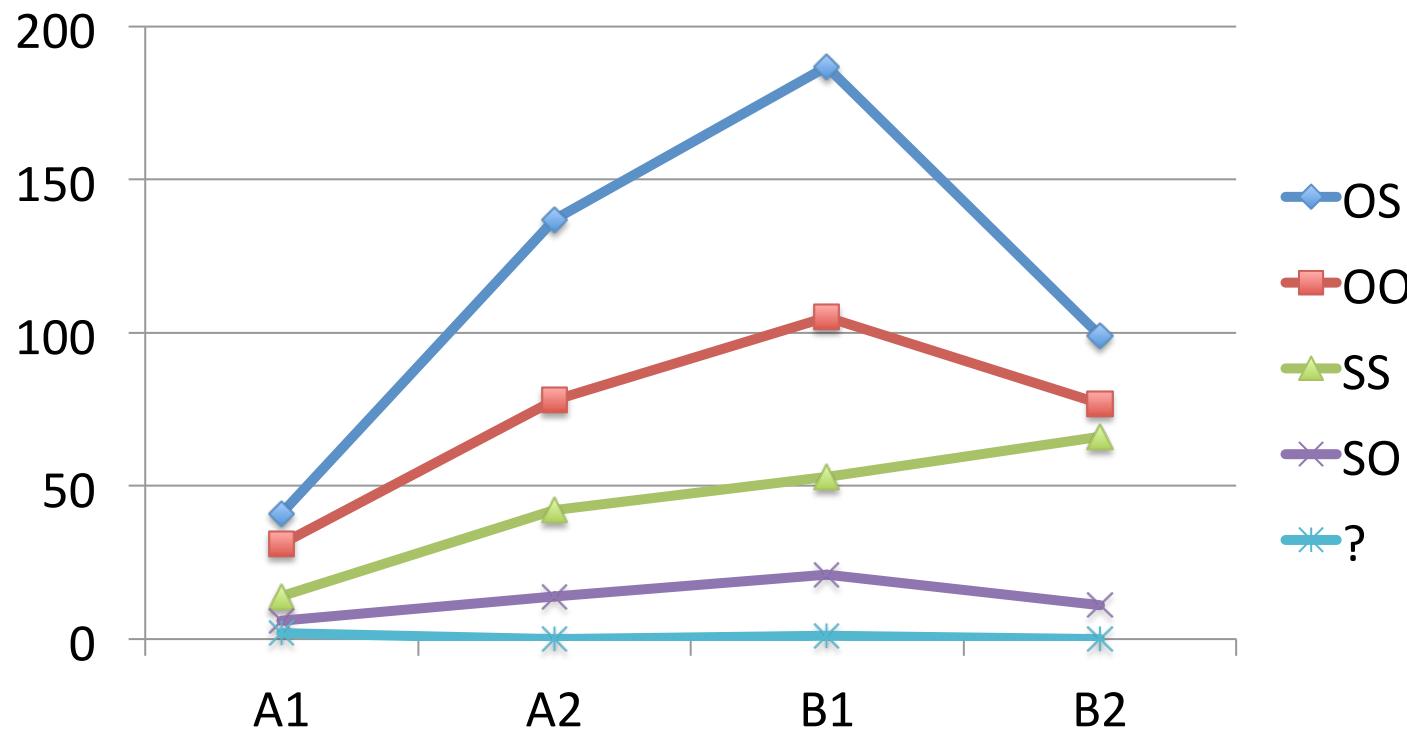
*Distributions of NPAH types across the CEFR levels
(Coursebook)*

SOHH (JEFLL)

Normalized frequencies of SOHH types across the CEFR levels (JEFLL)

SOHH	A1	A2	B1	B2	Total
OS	41	137	187	99	464
OO	31	78	105	77	291
SS	14	42	53	66	175
SO	6	14	21	11	52
unknown	2	0	1	0	3
Total	94	271	367	253	985

SOHH (JEFLL)



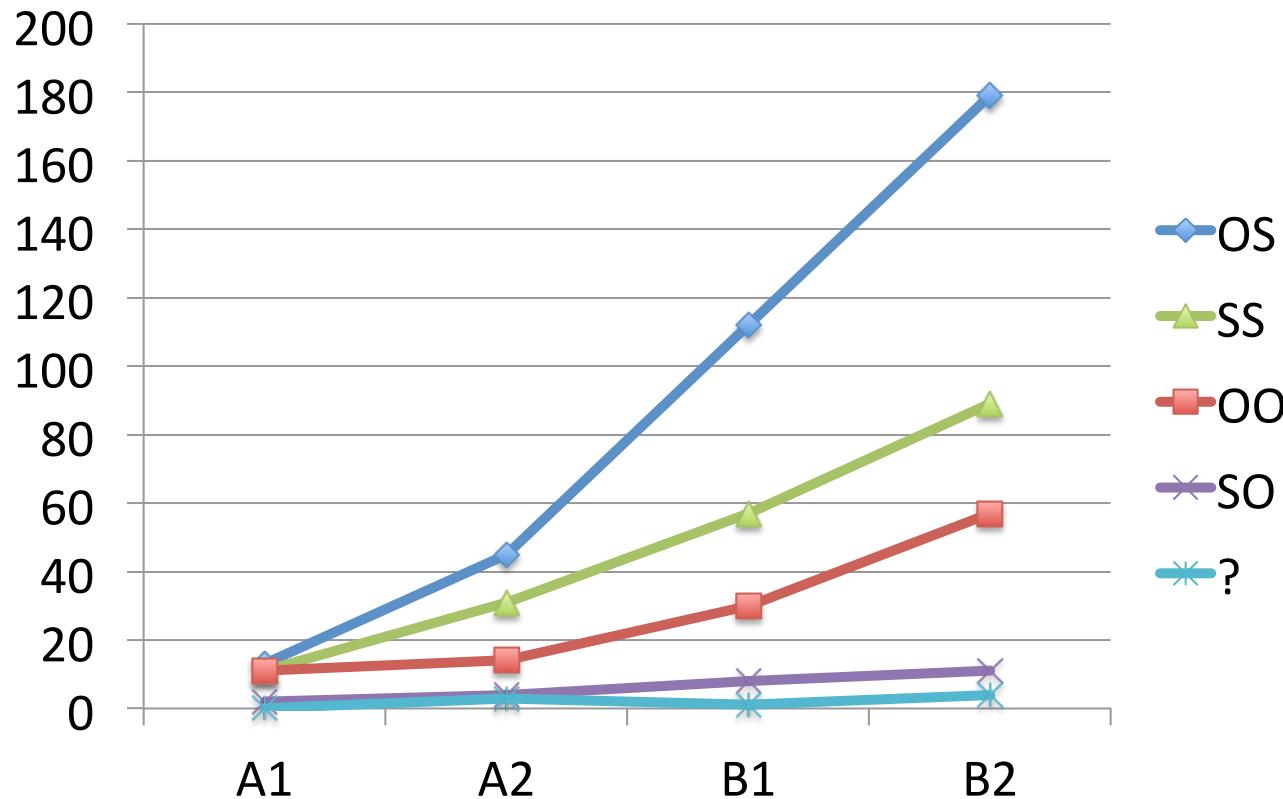
*Distributions of SOHH types across the CEFR levels
(JEFLL)*

SOHH (NICT)

Normalized frequencies of SOHH types across the CEFR levels (NICT JLE)

SOHH	A1	A2	B1	B2	Total
OS	13	45	112	179	349
OO	11	14	30	57	112
SS	11	31	57	89	188
SO	2	4	8	11	25
Unknown	0	3	1	4	8
Total	37	97	208	340	682

SOHH (NICT)



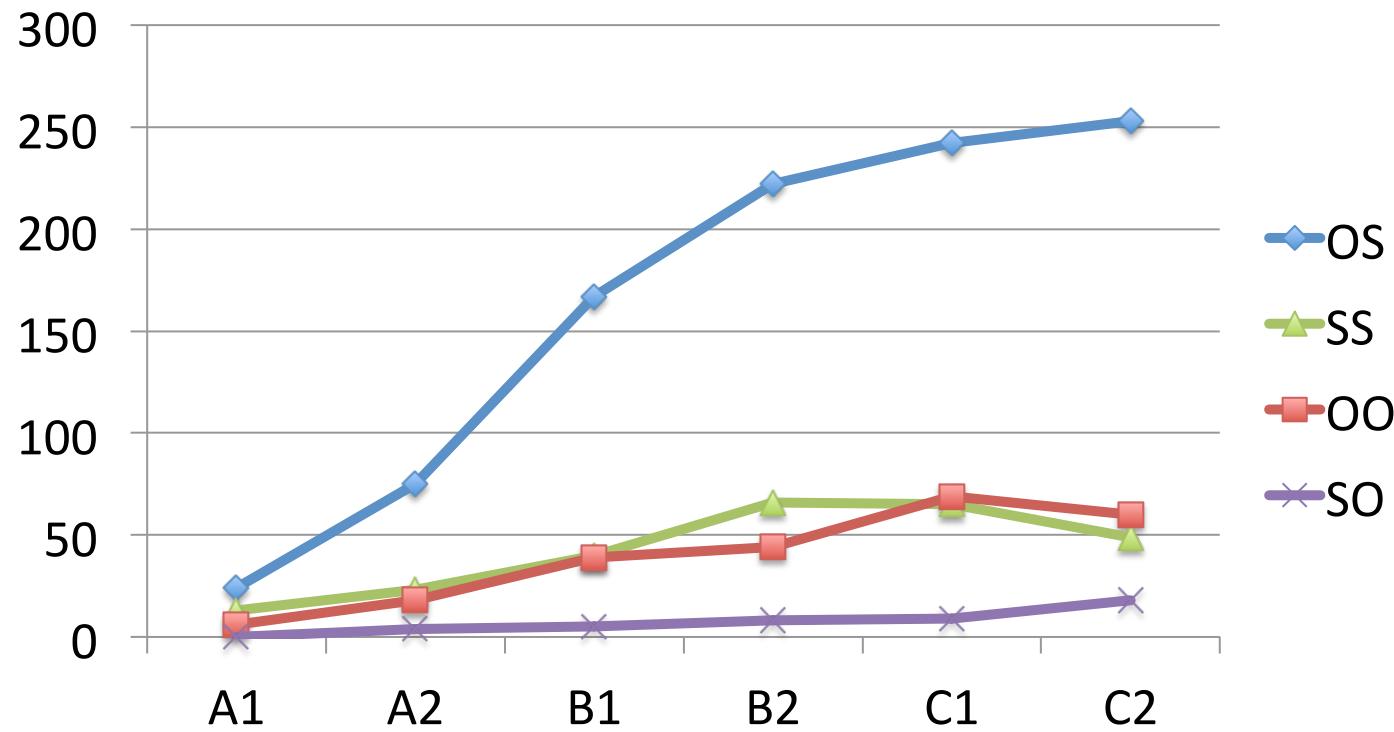
Distributions of SOHH types across the CEFR levels (NICT JLE)

SOHH (Coursebook)

*Normalized frequencies of SOHH types across the CEFR levels
(Coursebook)*

SOHH	A1	A2	B1	B2	C1	C2	Total
OS	24	75	167	222	242	253	983
OO	6	18	39	44	69	60	236
SS	13	23	40	66	65	49	256
SO	0	4	5	8	9	18	44
Total	43	120	251	340	385	380	1519

SOHH (Coursebook)

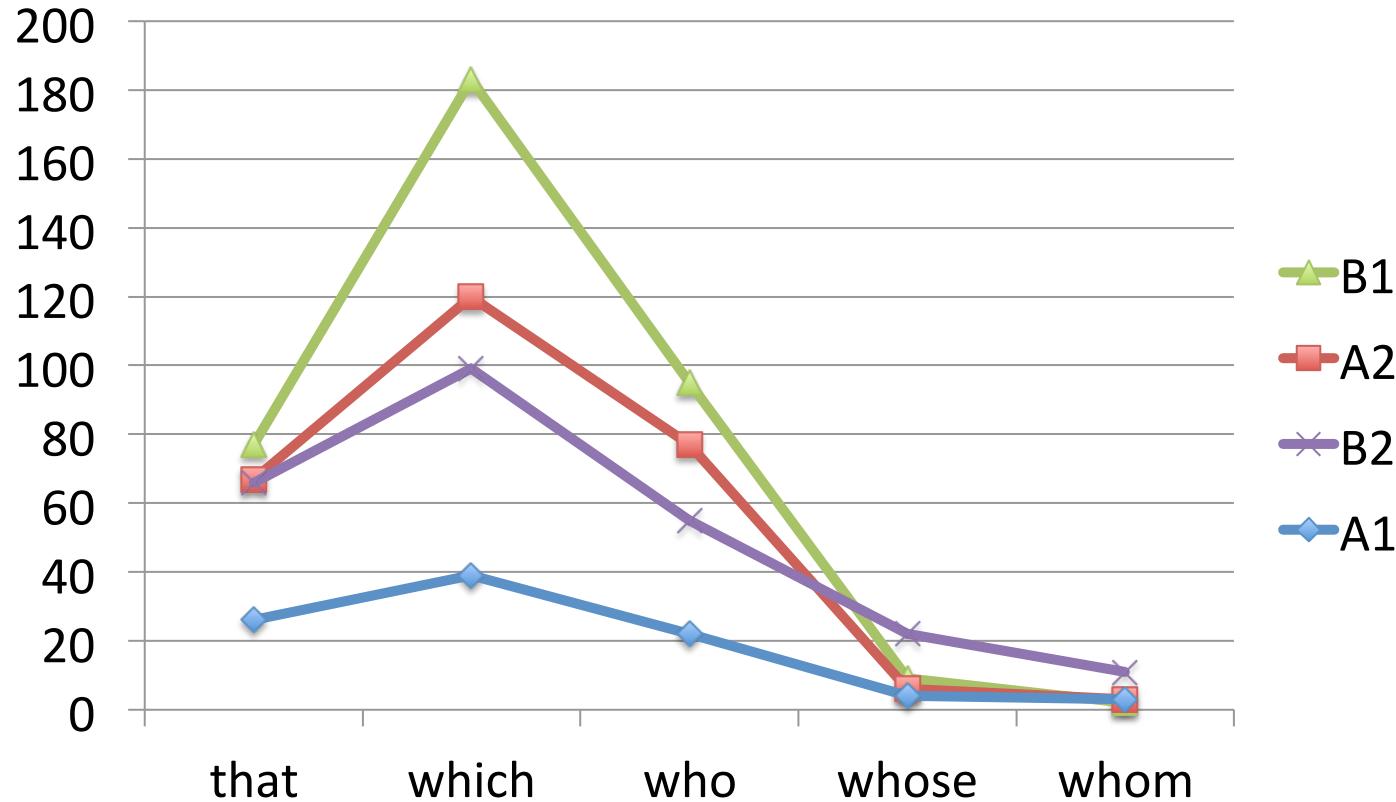


Distributions of SOHH types across the CEFR levels (Coursebook)

Q3

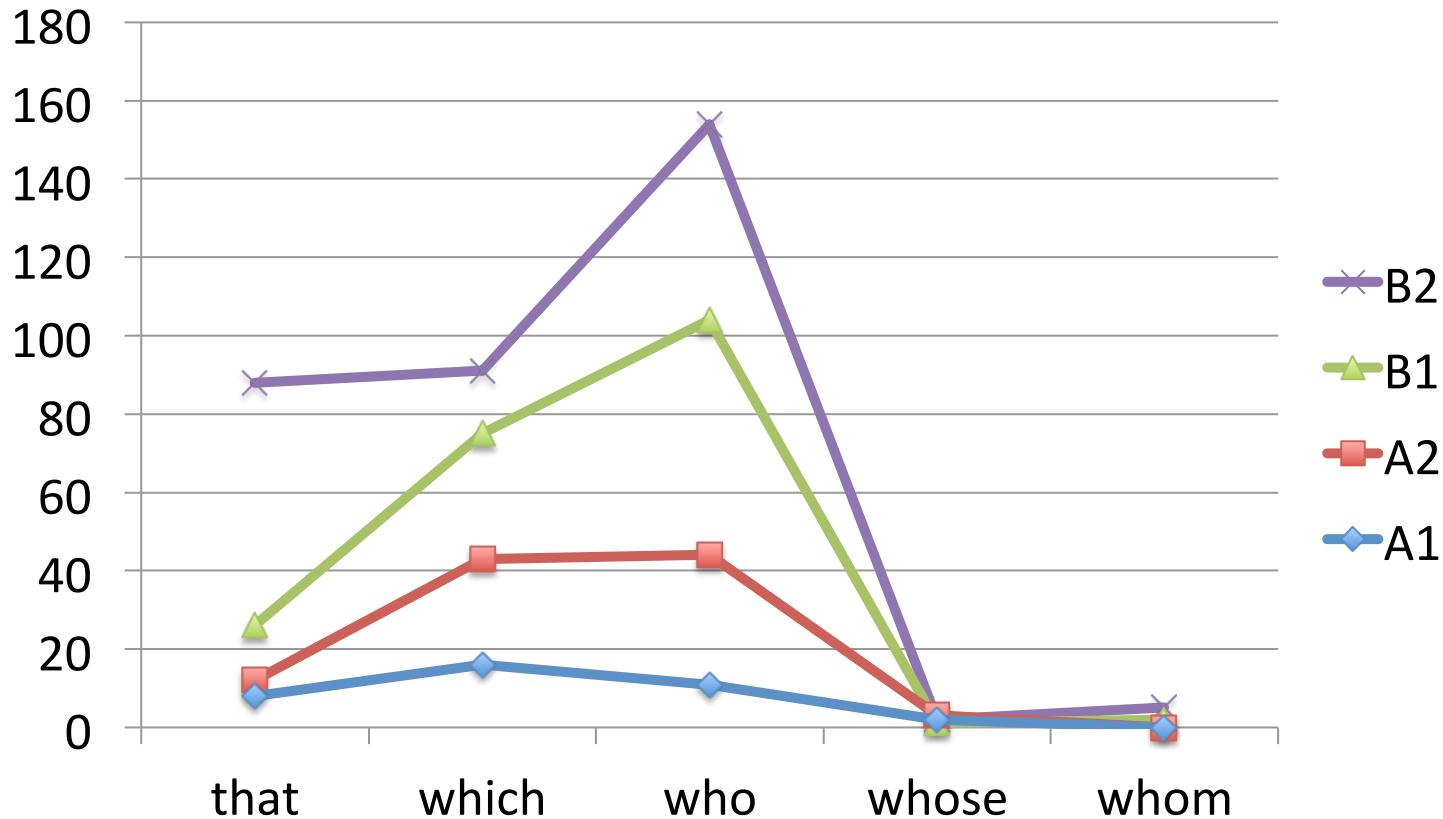
表層系、NPAH, SOHHの分類において、学習者の使用分布とコースブックの出現頻度の傾向に類似点・相違点はあるか。

表層形 (JEFLL)



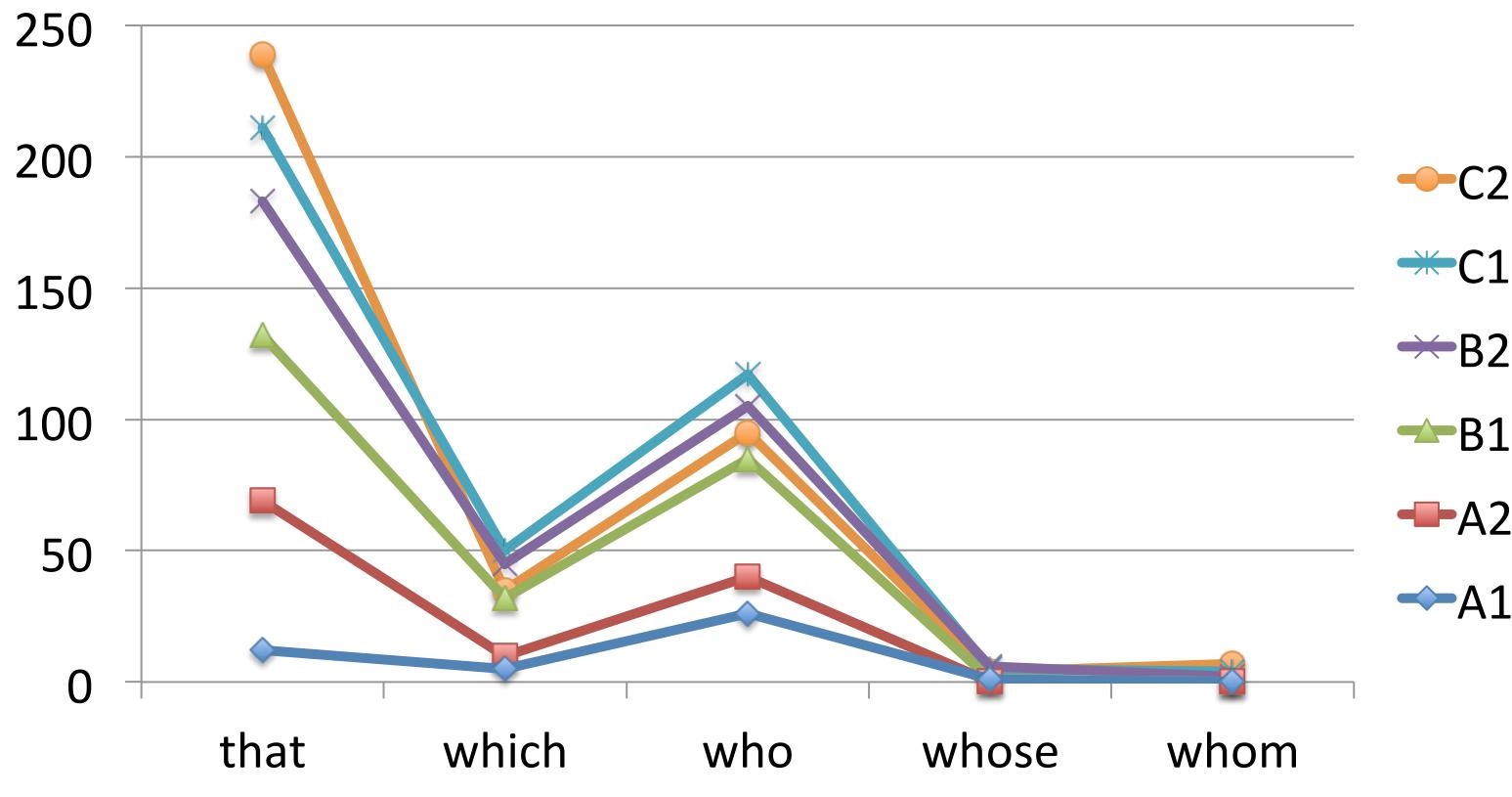
Normalized frequencies (100,000 per words) of each surface form (JEFLL)

表層形 (NiCT)



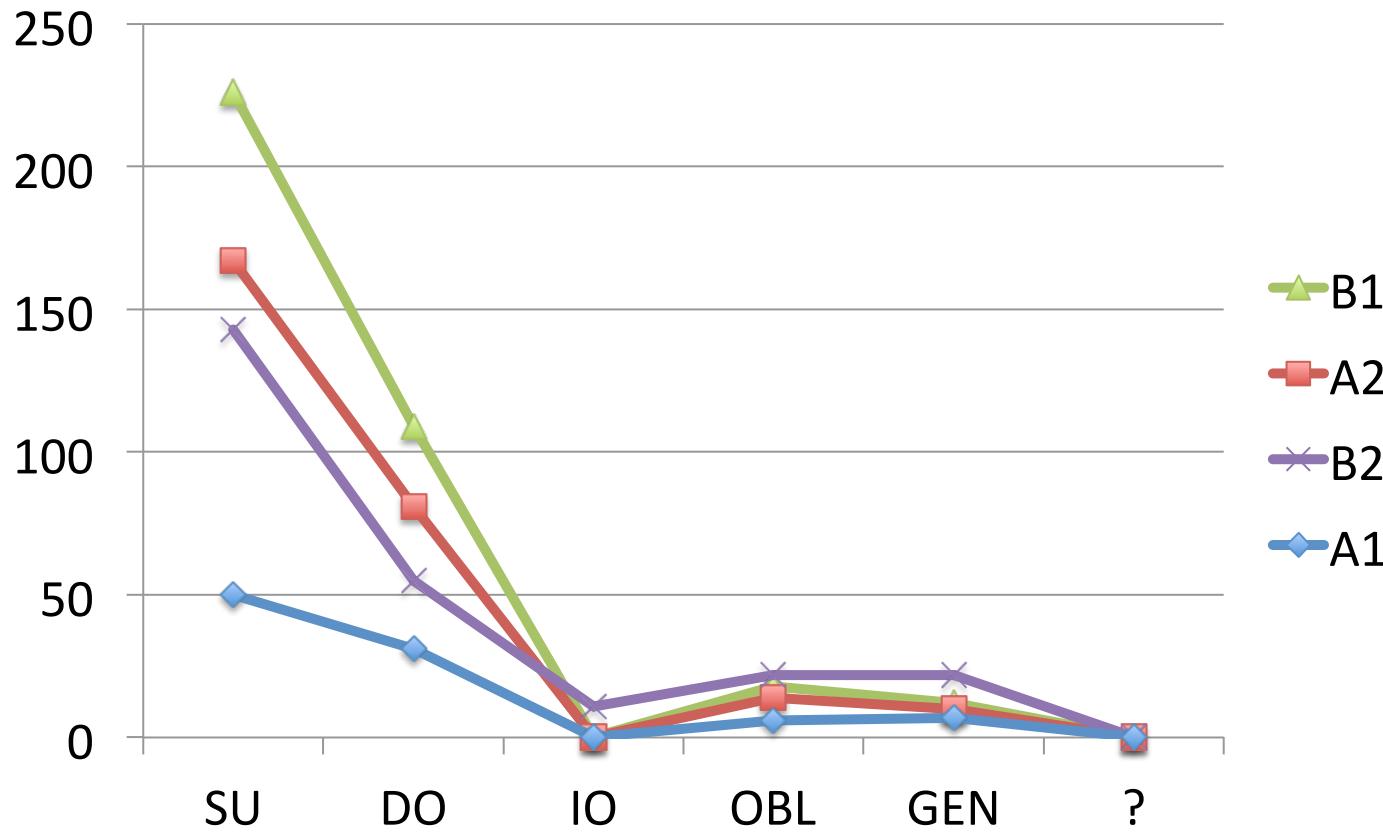
Normalized *frequencies* (100,000 per words) of each *surface form* (NiCT JLE)

表層形 (Coursebook)



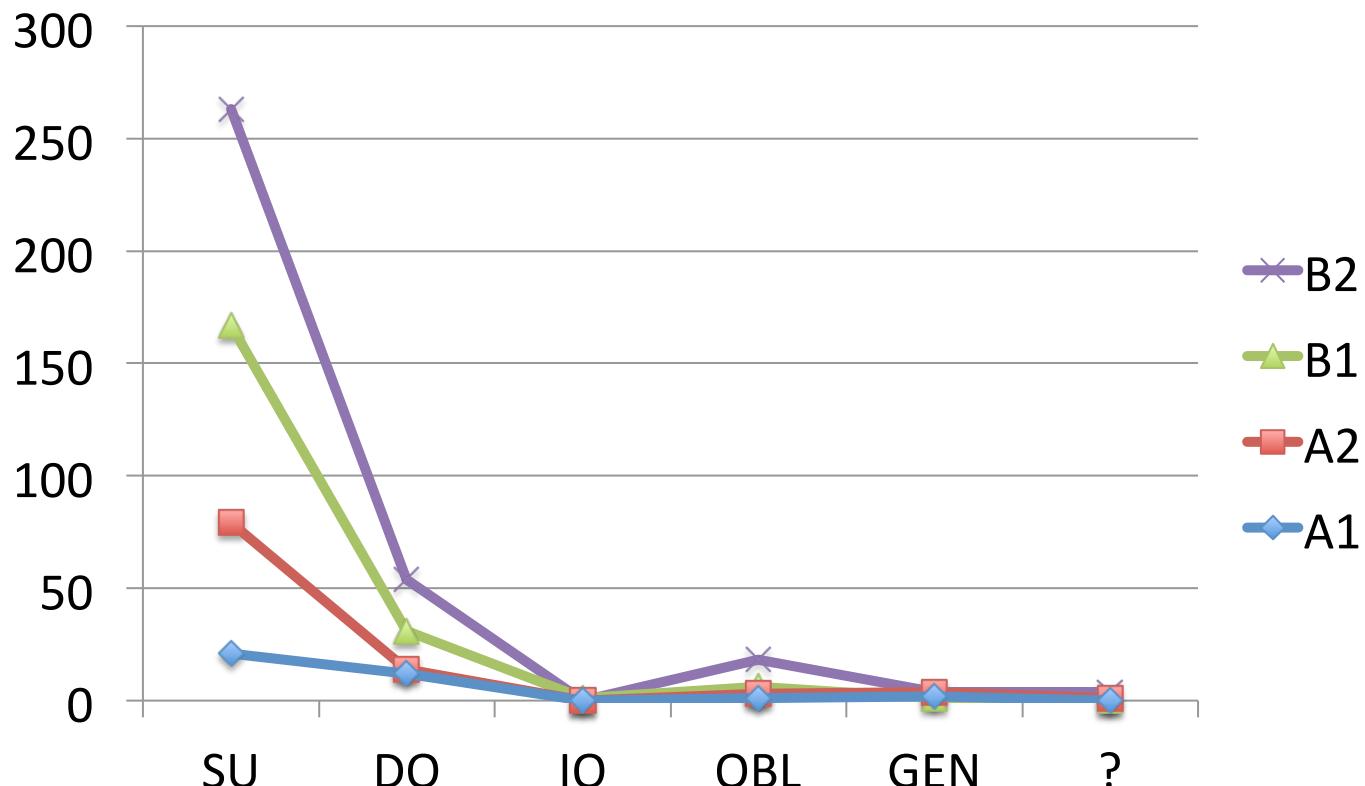
Normalized frequencies (100,000 per words) of each surface form (Coursebook)

NPAH (JEFLL)



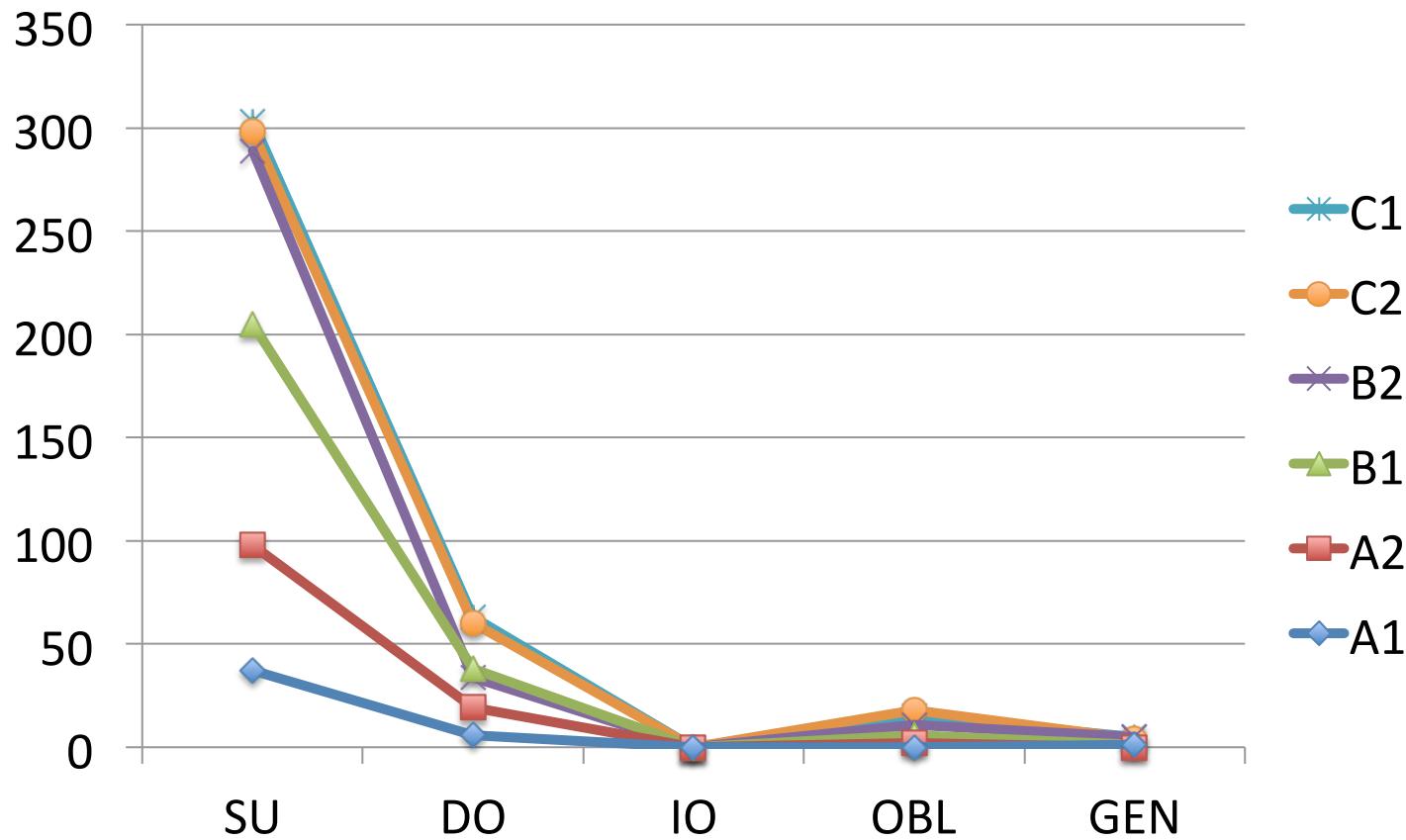
Distributions of each NPAH types (JEFLL)

NPAH (NICT)



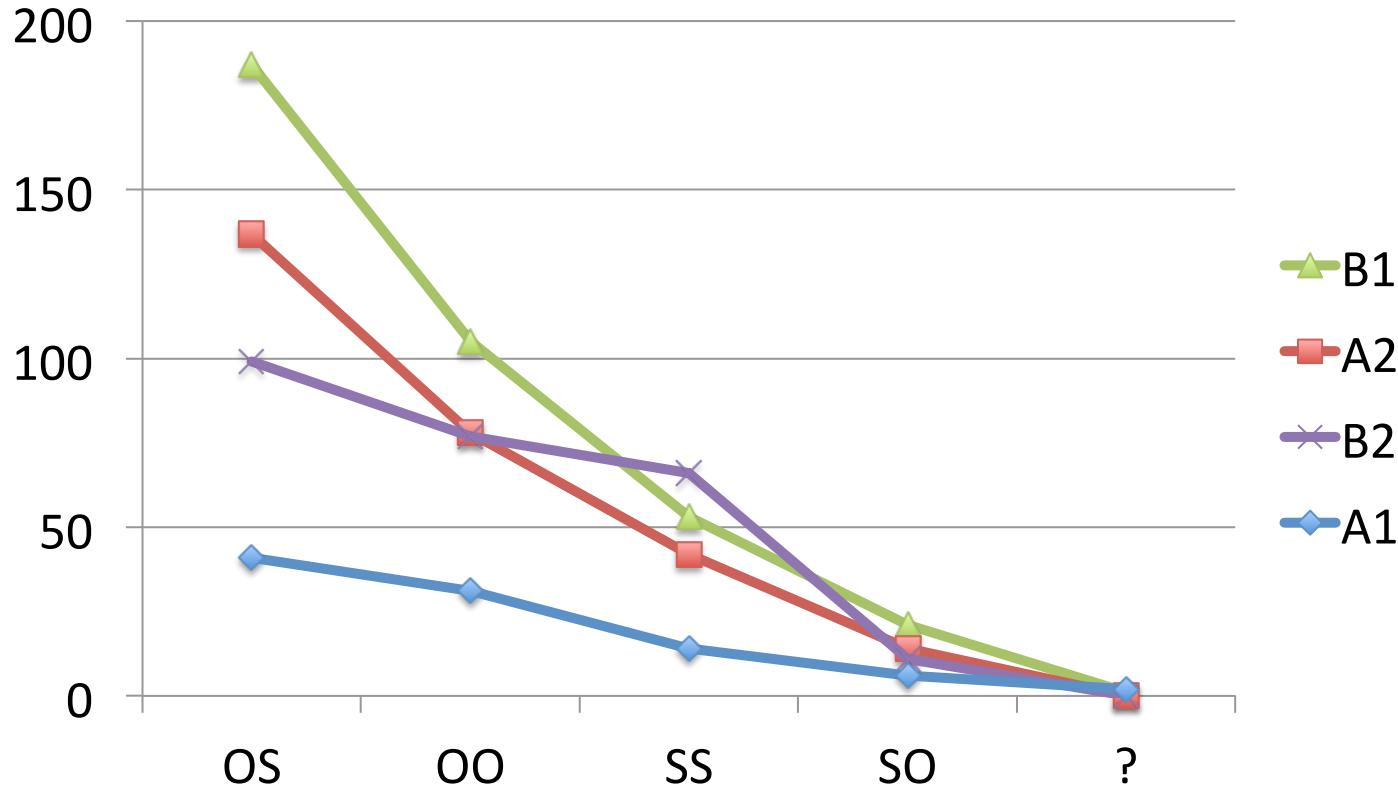
Distributions of NPAH types (NICT JLE)

NPAH (Coursebook)



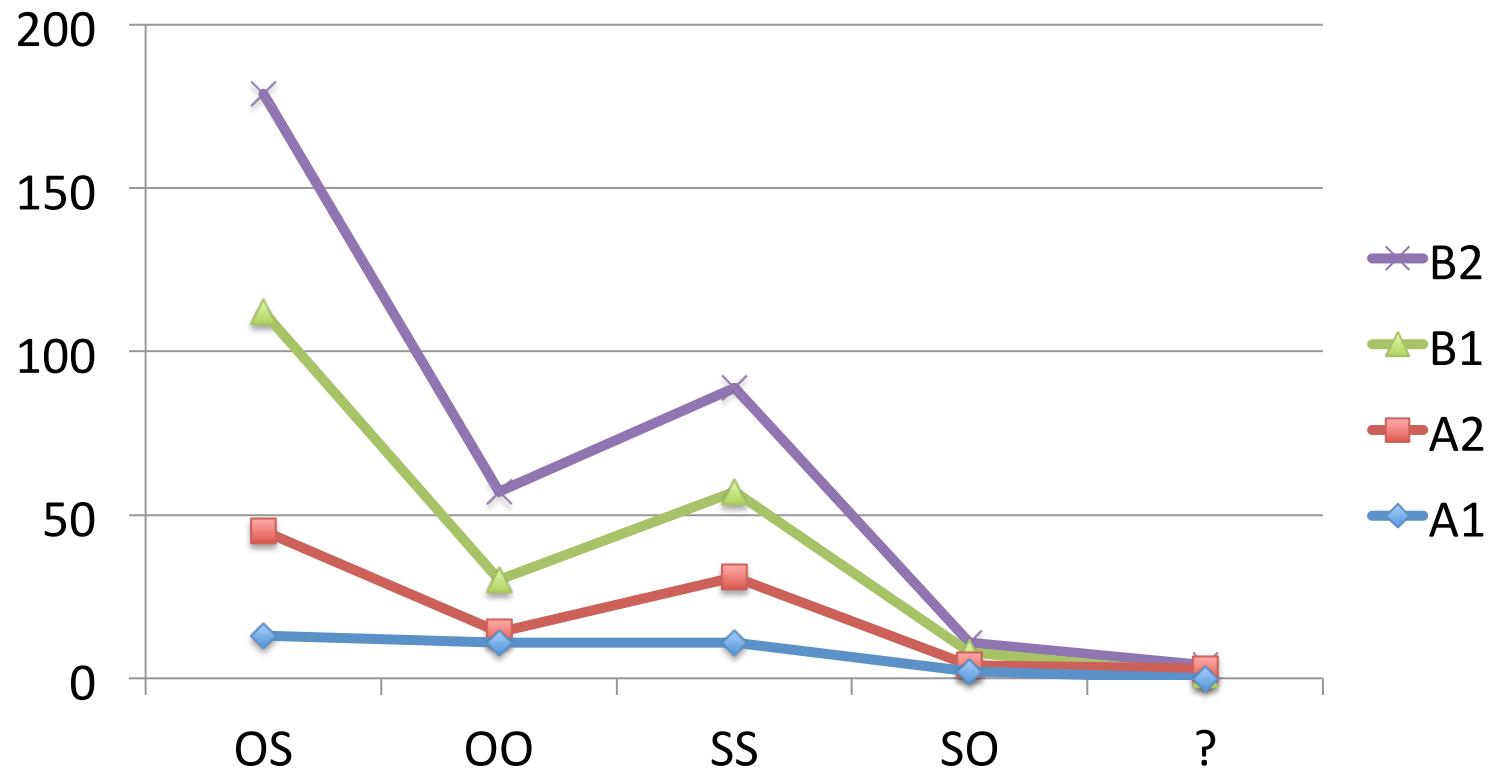
Distributions of NPAH types (Coursebook corpus)

SOHH (JEFLL)



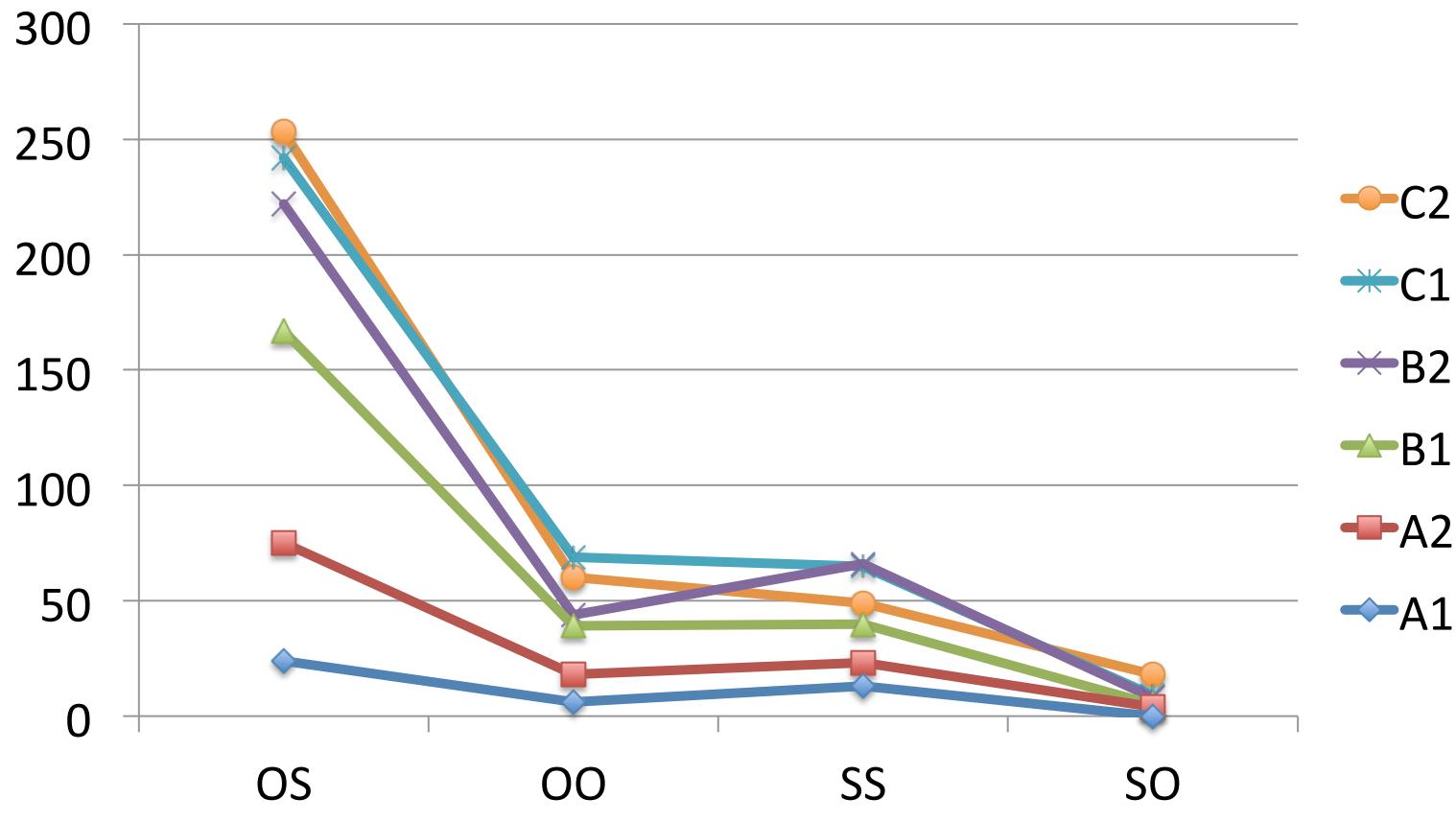
Distributions of SOHH types (JEFLL)

SOHH (NICT)



Distributions of SOHH types (NICT JLE)

SOHH(Coursebook)



Distributions of SOHH types (Coursebook)

Q4

各レベルにどのようなエラーがみられるか。書き言葉と話し言葉の間に違いは見られるか。

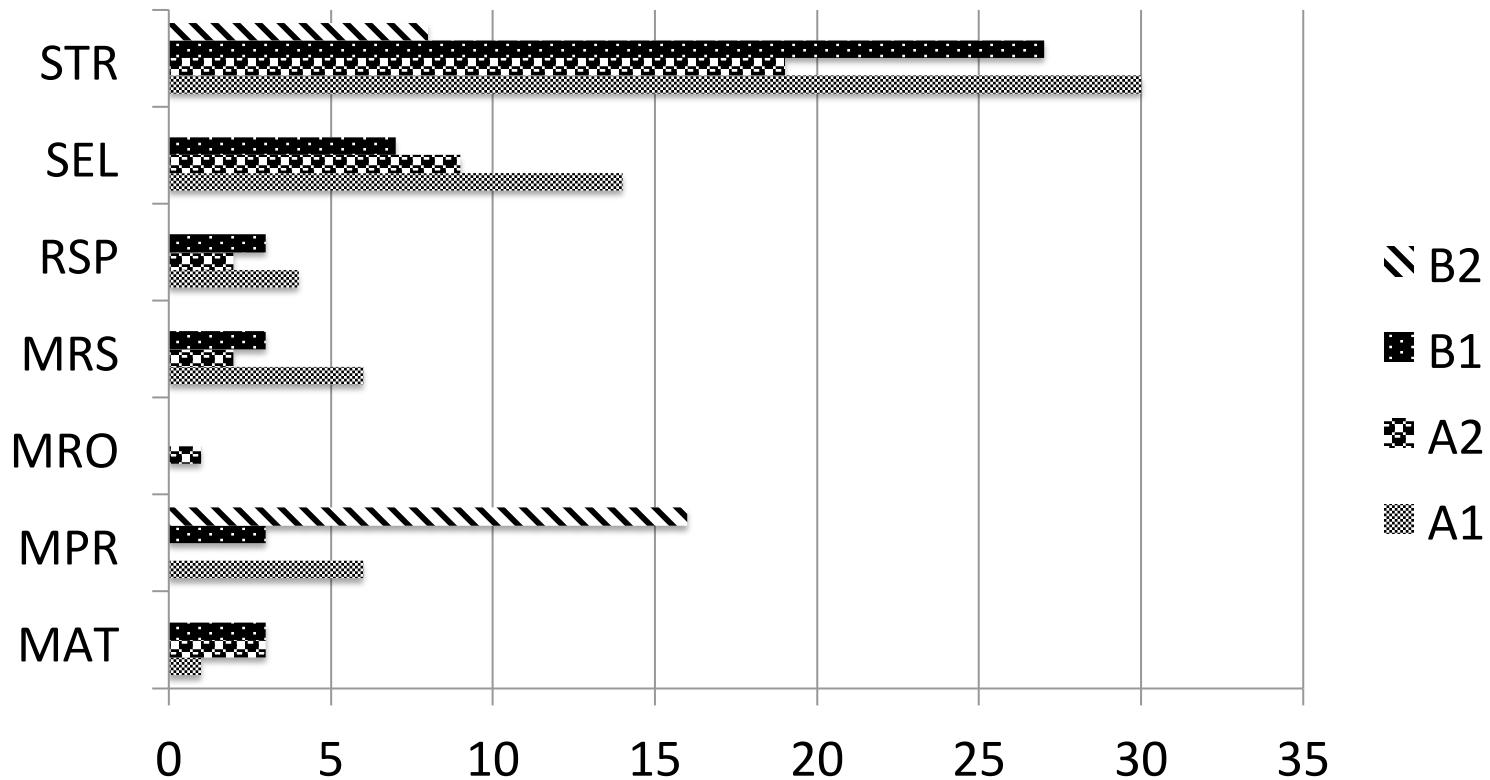
Numbers of correct use and errors with percentages (JEFL)

JEFL	RC sentences	%
Correct use	1401	77.53%
Errors	406	22.47%
Total RC sentences	1807	100.00%

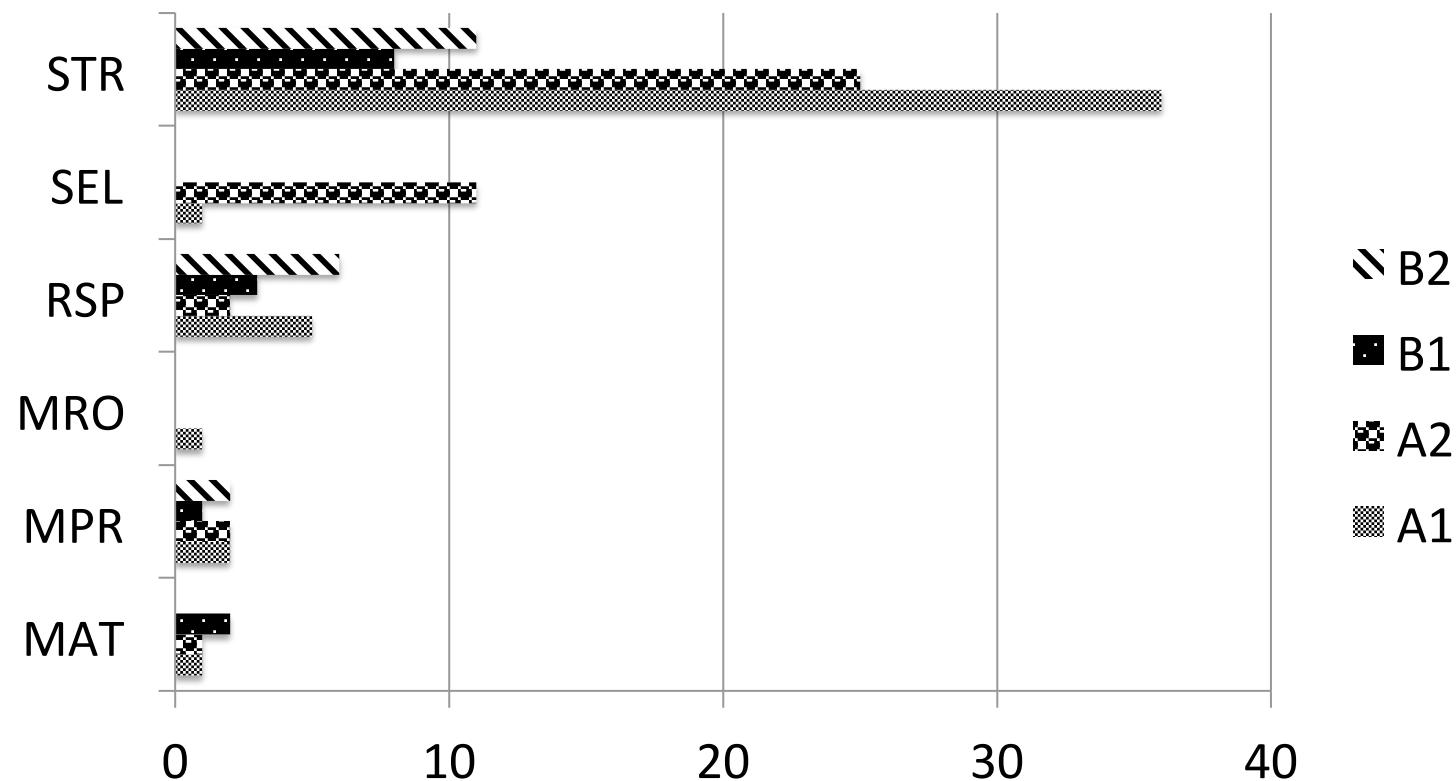
Numbers of correct use and errors with percentages (NICT JLE)

NICT	RC sentences	%
Correct use	1188	87.35%
Errors	172	12.65%
Total RC sentences	1360	100.00%

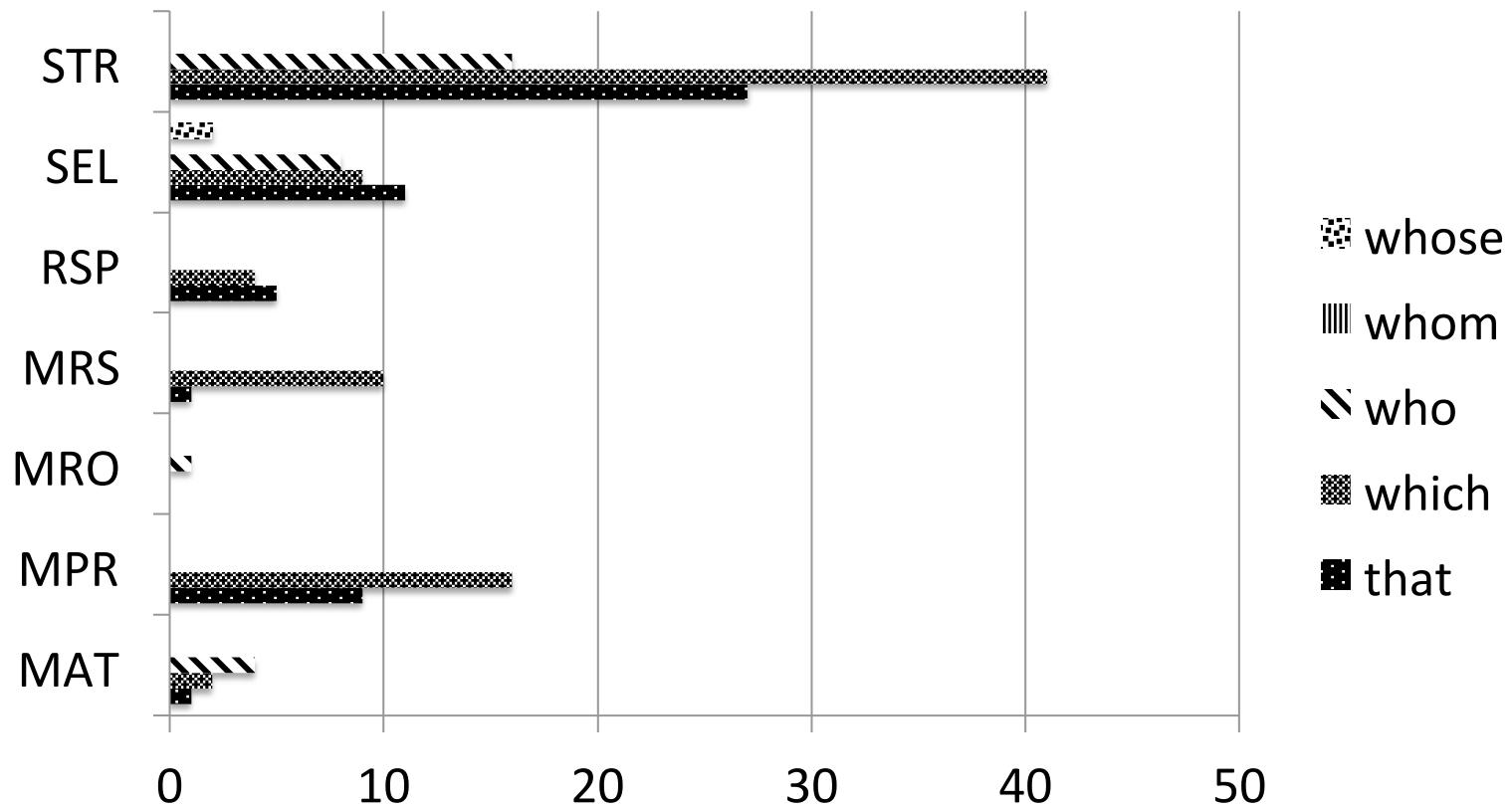
Sampled frequencies along the CEFR levels (JEFLL)



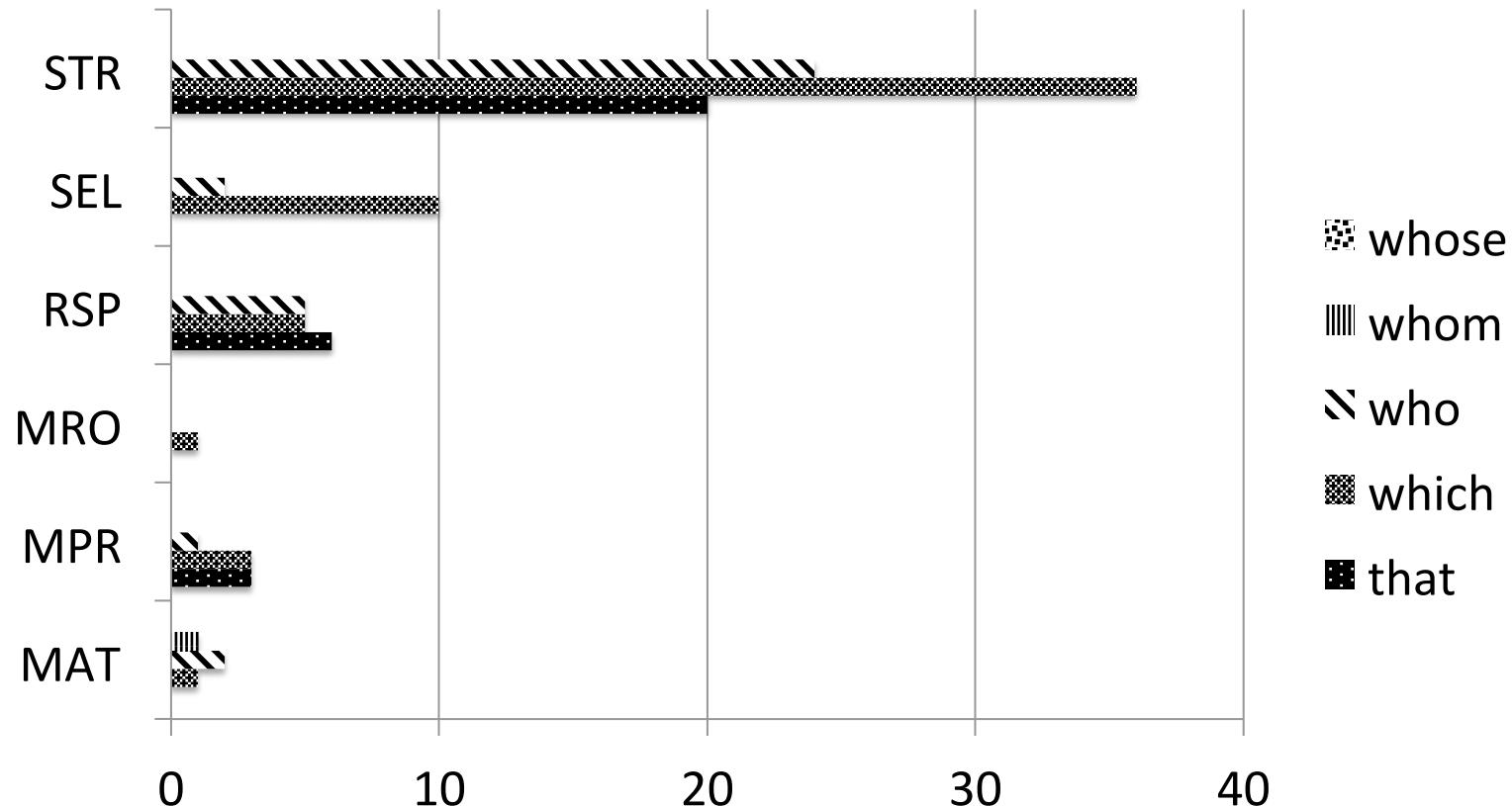
Sampled frequencies along the CEFR levels (NICT JLE)



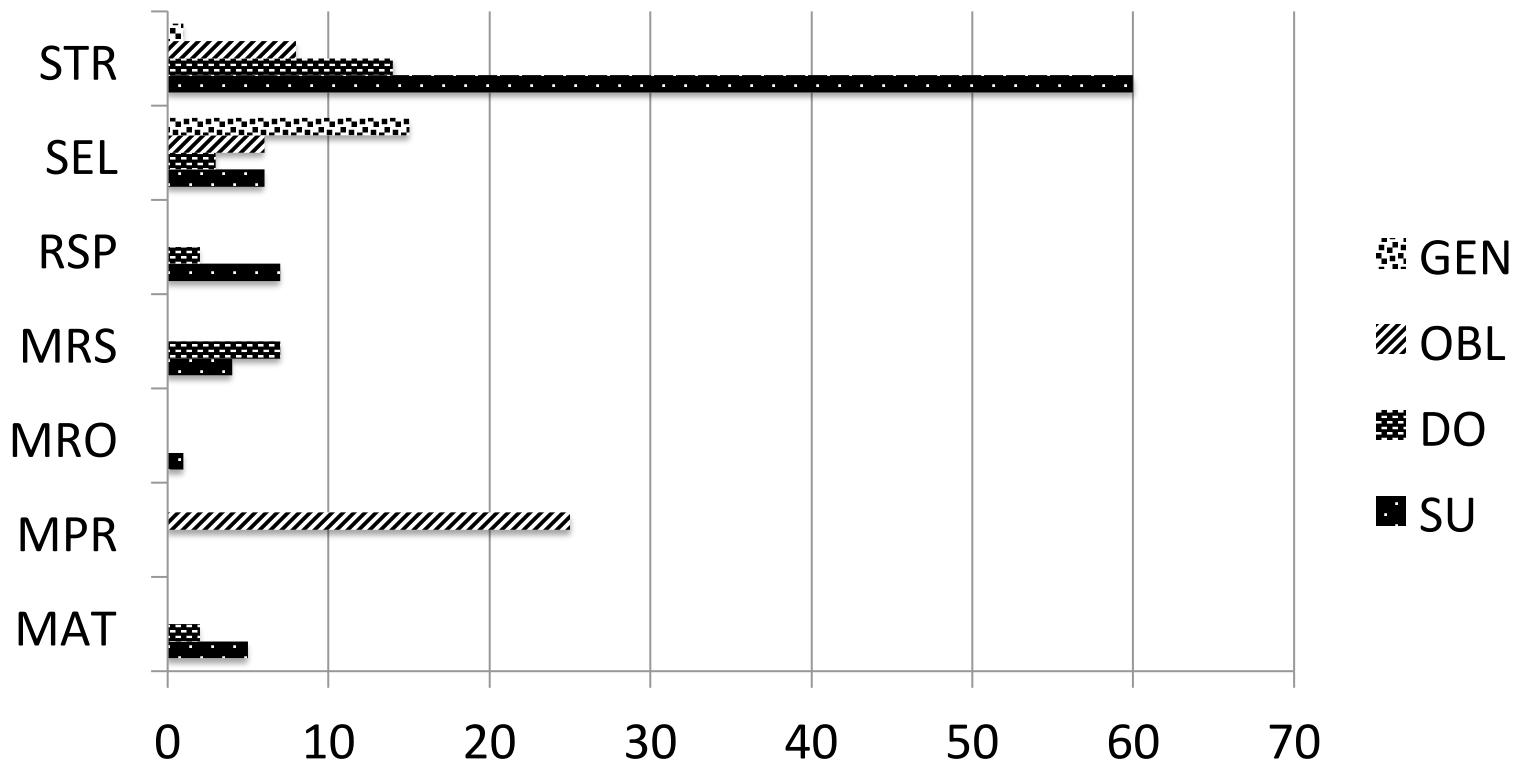
Sampled frequencies in each surface form (JEFLL)



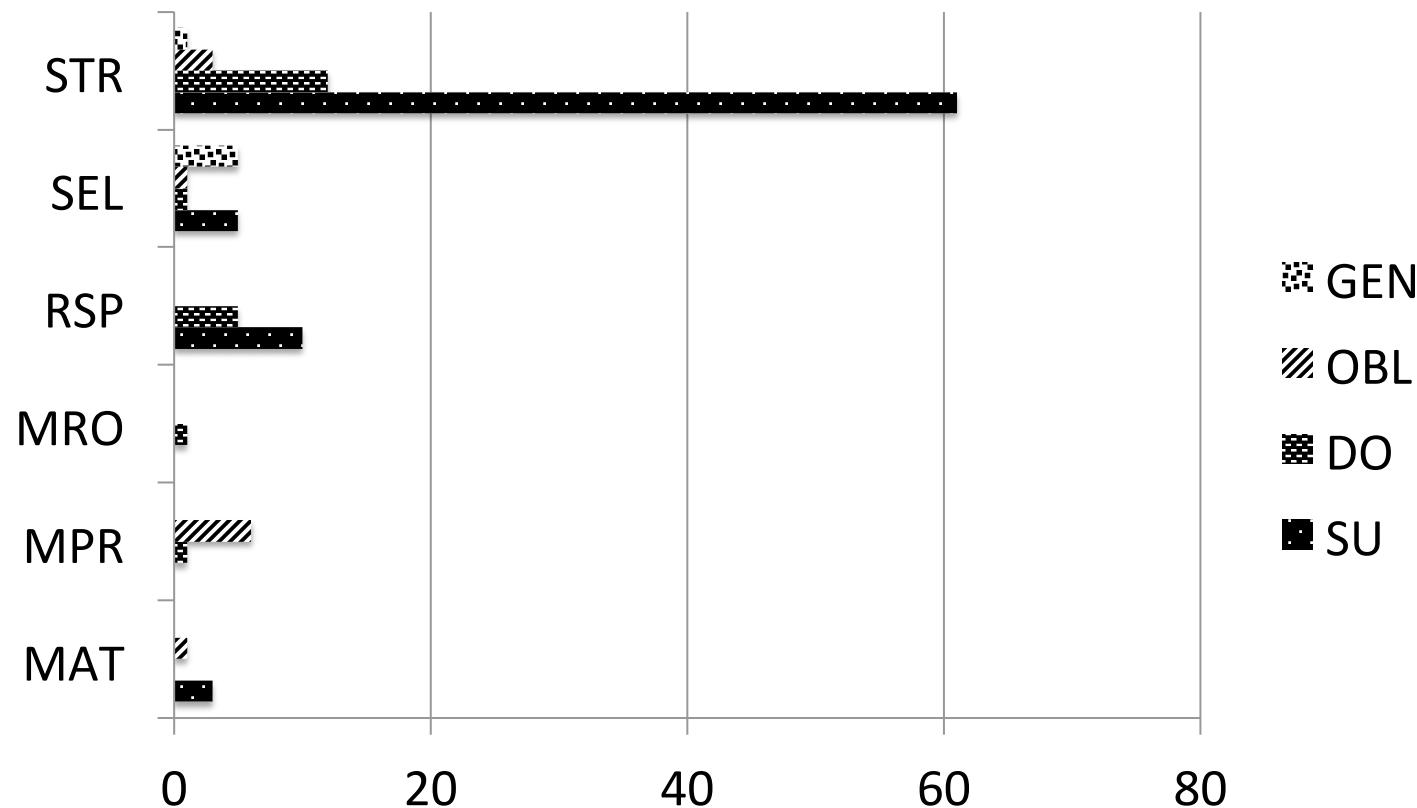
Sampled frequencies in each surface form (NICT JLE)



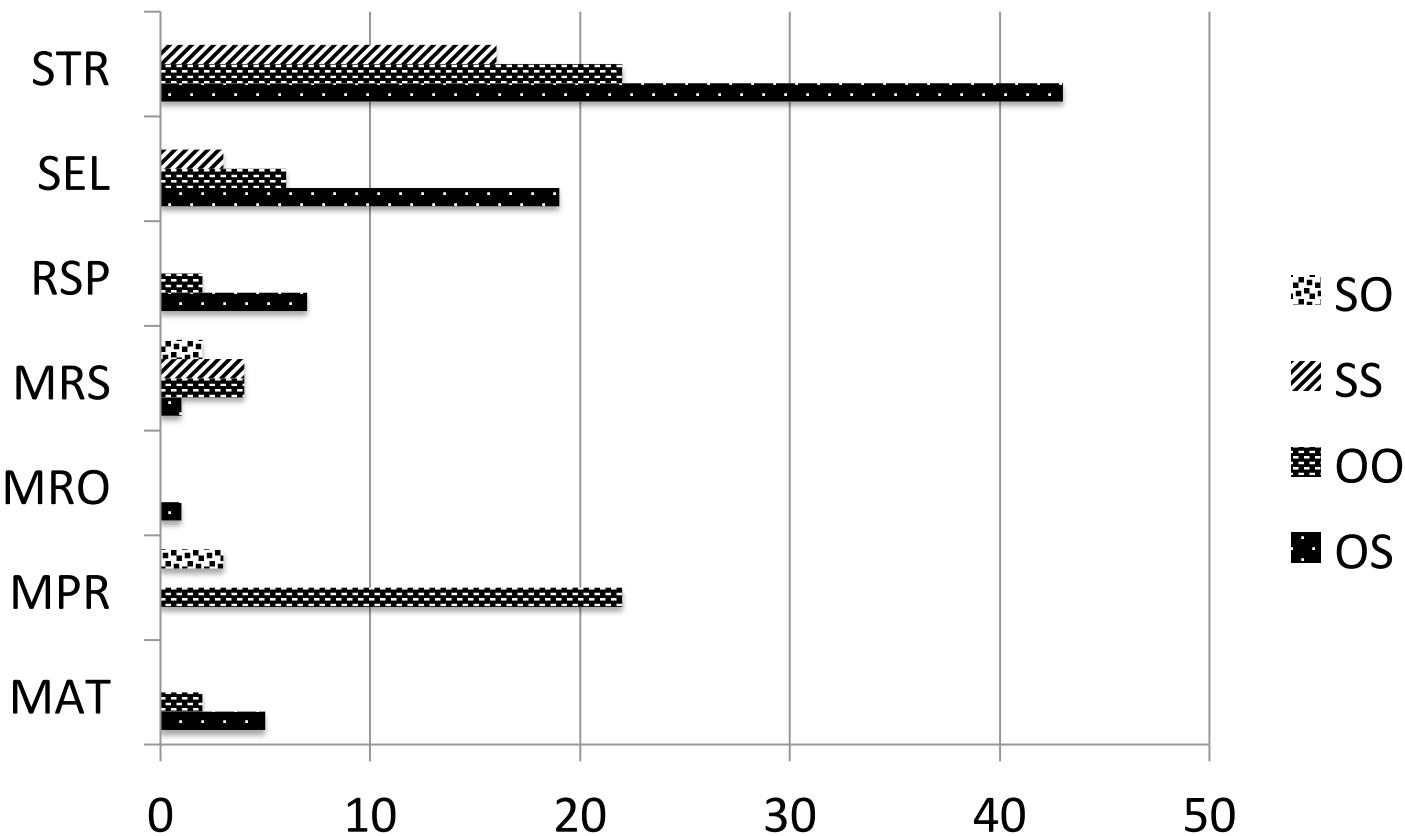
Sampled frequencies in each NPAH types (JEFLL)



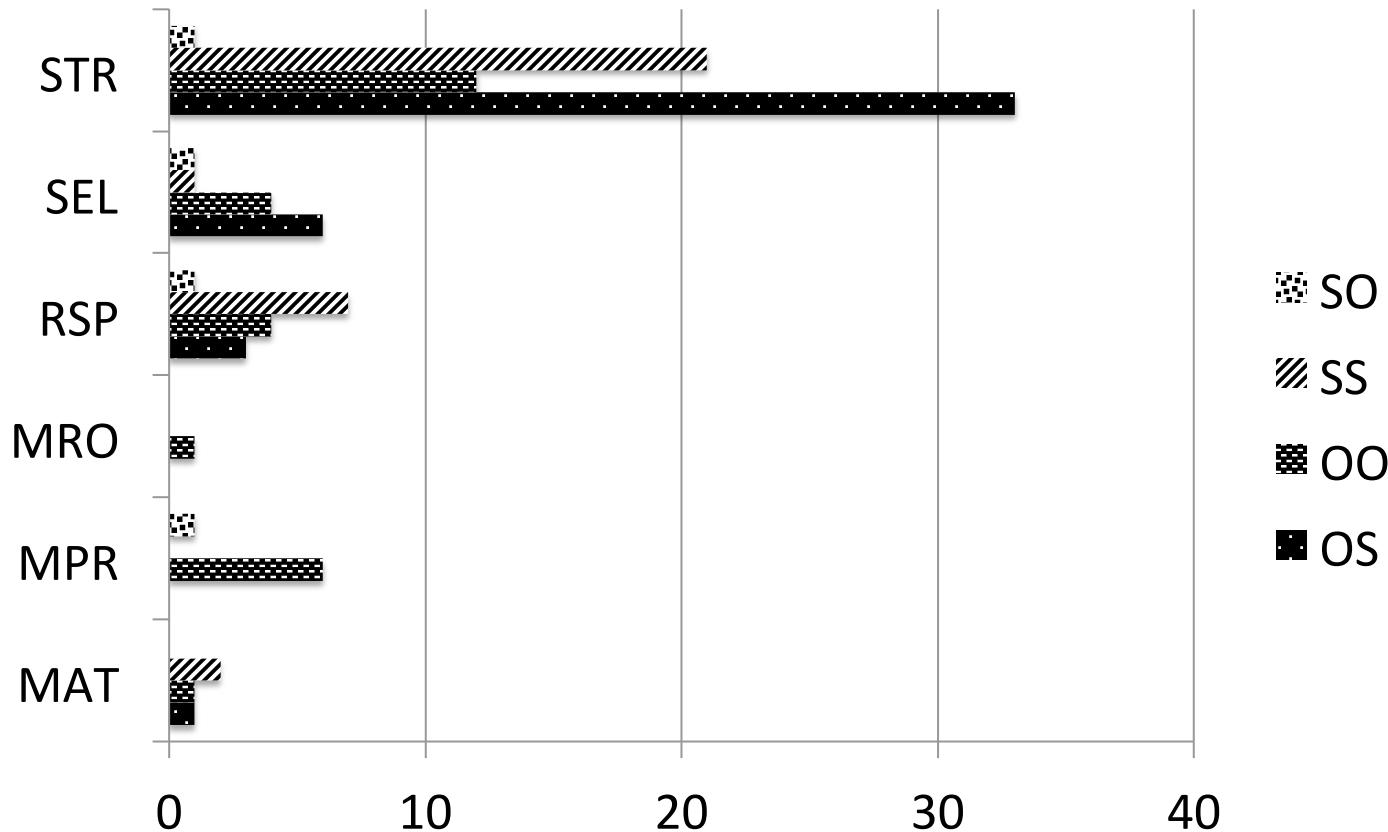
Sampled frequencies in each NPAH types (NICT)



Sampled error frequencies in each SOHH type (JEFL)



Sampled error frequencies in each SOHH type (NICT JLE)



結果と考察

- RQ1: 関係詞の使用と CEFRレベル
 - 関係詞の使用は特にthat, who, whichに関してはCEFRレベルが上がると増加し、基準特性として機能すると思われる。
 - whose, whomは低頻度だが、上位レベル(B2以上)とそれ以外を分けるのには有効
- RQ2: 関係詞タイプとCEFRレベル
 - NPAH, SOHHもほぼ仮説の順序に等しかった
 - 間接目的語(IO)のみがコースブックでも学習者データでも仮説より順序が低かった
- RQ3: CB vs. 学習者データ
 - CB: thatが多く、全体に口語体の関係詞の使用を反映
 - 日本の教科書に比べるとCBはthat重視、whose/whomはかなり扱いが薄く、例文もBレベルにならないとでてこない。
 - A levels / B1 / B2-C2
 - JEFLL: whichが多く、書き言葉の傾向により似ている
 - A1 / B2
 - NICT: whoが多く、thatが会話なのに過少使用気味
 - CBで扱われるようなthatの用法があまり身についていない可能性？
 - タスク・トピックの影響
 - A levels / B levels
- RQ4: エラーの特徴、書き言葉/話し言葉での差異
 - エラーの個別のタイプに関しては、選択エラー、代名詞残留、構造エラーなどが顕著
 - 選択&構造エラーはレベルが上がると比較的減少し、代名詞残留、前置詞欠落はあまり減らない
 - エラータイプ別の気づきを与えるような指導が必要かもしれない

課題

- Bレベル以上のデータを含める
- 省略形を含める
- 日本の教科書との関連性
- 母語話者との比較
- トピックの影響
- タグ付けの信頼性
- 回避の割合
- 母語の影響

参考文献

- Hamilton, R. 1994. 'Is implicational generalization unidirectional and maximal? Evidence from relativization instruction in a second language'. *Language Learning* 44: 123-57
- Hawkins, J. A., 2009. Cambridge / UCLES-RCEAL Research Projects. Internal report.
- Hawkins, J. A., & Filipovic, L. 2012. *English Profile Studies, Criterial Features in L2 English, Specifying the Reference Levels of the Common European Framework*. Cambridge: Cambridge University Press.
- Izumi, S. 2003. 'Processing difficulty in comprehension and production of relative clause by learners of English as a second language'. *Language Learning* 53:285-323
- Keenan, E. L., and Comrie, B. 1977. 'Noun phrase accessibility and universal grammar'. *Linguistic Inquiry*, 8(1), 63-99.
- Shirai, Y. and Ozeki, 2007. 'Introduction'. *Studies in Second Language Acquisition* 29: 155-67.
- Takahashi, Y. & Tono, Y. 2014. A learner corpus-based study on relative clause constructions as criterial features for the CEFR levels. Poster presented at TALC2014, Lancaster University.
- Zobl, H. 1989. 'Canonical typological structures and ergativity in English L2 acquisition' in S. Gass and J. Schachter (eds.): *Linguistic Perspectives on Second Language Acquisition*. Cambridge: Cambridge University Press
- 投野由紀夫 (編著) (2007).『日本人中高生1万人の英語コーパス JEFLLコーパス 中高生が書く英文の実態とその分析』東京: 小学館