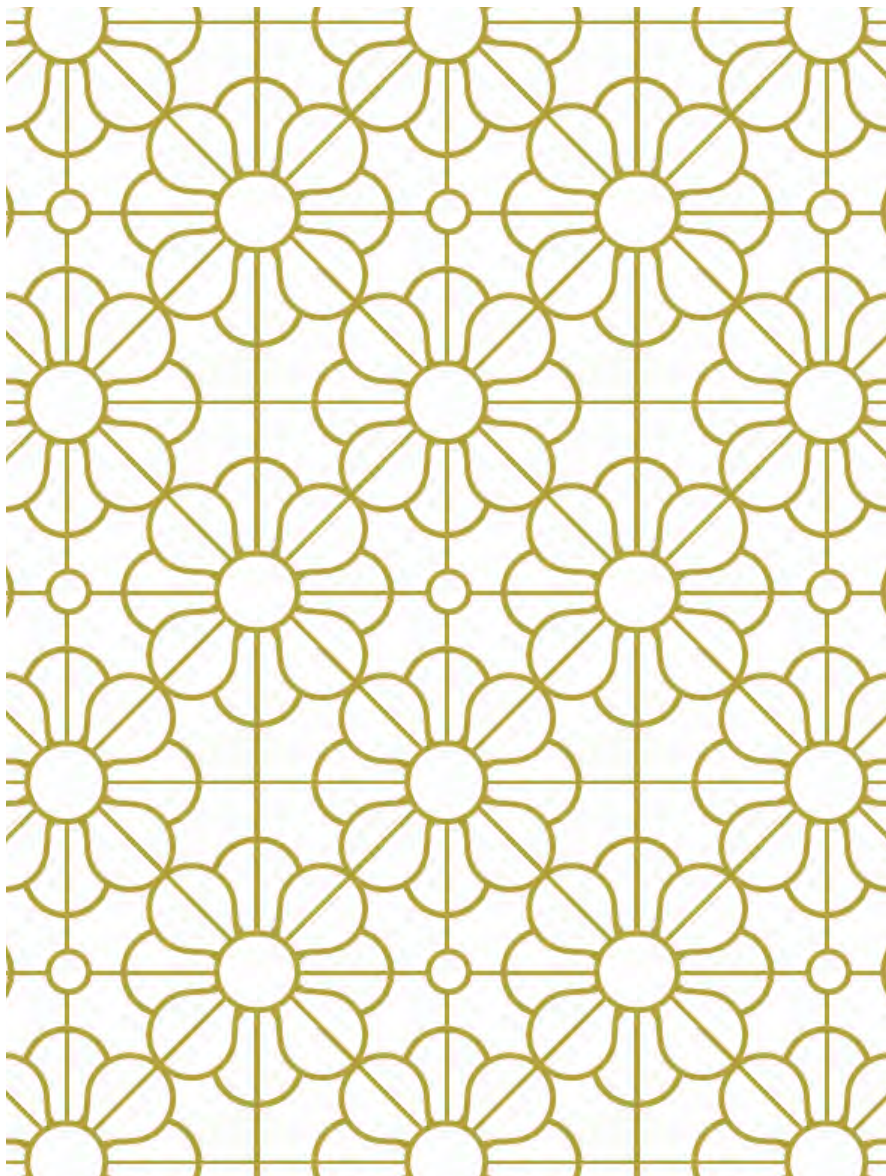




談話標識抽出手法の検討 ーTUFS2010を例にー

秋廣尚恵・國末薫

東京外国語大学・東京外国語大学特別研究員

2026年6月14日 ワークショップ 母語話者言語・学習者言語研究

発表の流れ

はじめに

先行研究

調査方法

分析結果

考察

結論・今後の課題

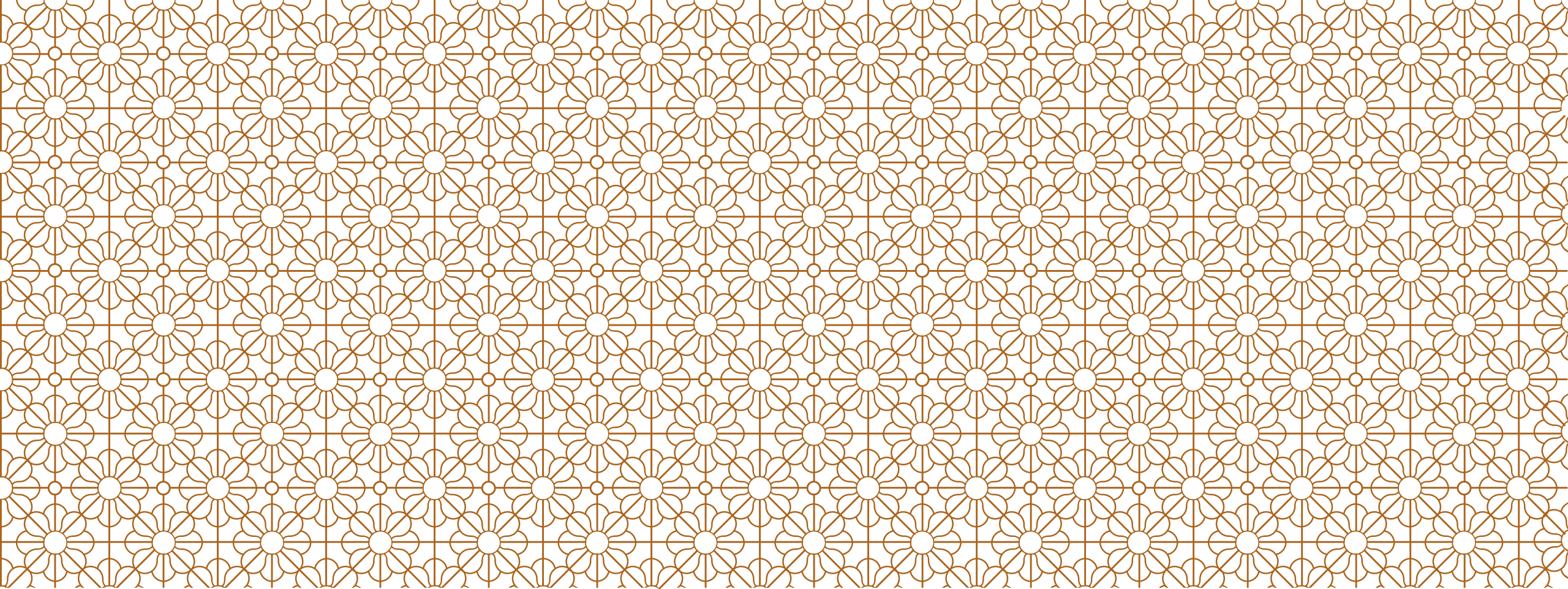
はじめに

■ 本発表の目的

- ・ 2025年度実施の談話標識の調査内容報告
- ・ 談話標識抽出手法についての検討

■ 調査の概要

- ・ TUFS 2010 のデータより、談話標識を抽出し分析
- ⇒本発表では、2500例についての一次調査の結果を報告



先行研究



談話標識研究の始まり

【背景】

- 談話構造の解明への関心増大。
- 話し言葉の音声/録画データと転写やアノテーションのあるコーパスデータの拡充
- コーパス検索ツールやデータ共有の進歩
- 話し言葉や談話を分析対象とした様々なアプローチや理論的枠組みの提唱。

【談話標識の研究の始まり】

- Shiffrin (1987) の研究に始まり、1990年以降、35年にわたり様々な言語で研究が盛んにおこなわれてきた。
- Schifffrin, Fraser, Redeker, Hansenなどの前期（1990年代から2010年代）の研究が提唱した用語や概念はいまだに整理しなおされつつ、多くの研究で参照されている。
- 英語圏の様々な研究、とりわけ用法基盤モデル、認知言語学的観点からの研究（Traugottの文法化理論や関連性理論）の潮流も大きな影響を与えている。

談話標識研究の現状

□ 複数の視点からの分析方法や理論的枠組みが併存している。

- Shiffrin, Fraserの語用論的分析
- Hansen の意味語用論的分析
- Redeker の談話標識の領域と機能の定義
- Anscombe, Ducrot などの論証的言語学理論
- Paillardらキュリオリ派の発話理論研究
- Kerbrat-Orrechioni らのやり取りの言語学
- 関連性理論 (Pons Bordería 2008)
- マクロ統語論 (Blanche-Benveniste, Berrendonner)
- 主観性、間主観性といったモダリティ研究
- 文法化理論 (Traugott, Degand & Evers-Vermeul など)
- Dostie, Beeching らの社会言語学的変異の研究など

談話標識研究の発展

【最近の動向】

- 言語間比較研究 (Heine 他 2021)
- 学習者の談話標識習得 (Beechingなど) * ケーススタディは多く行われている。
- 実験的研究による認知科学的研究 (Broisson & Degand など)
- 自動言語処理による半自動的アノテーションの開発 (Crible & Degand 2019)
- 社会言語学的変異の研究 (Dostie, Beechingなど)

→ 言語学の様々な隣接分野や技術分野を巻き込みつつ、研究が発展している状況にある。

問題点

❖ 「発話、発話行為といった単位を結びつけ談話を構築するために、手続き的意味を伴って使われる機能的表現」

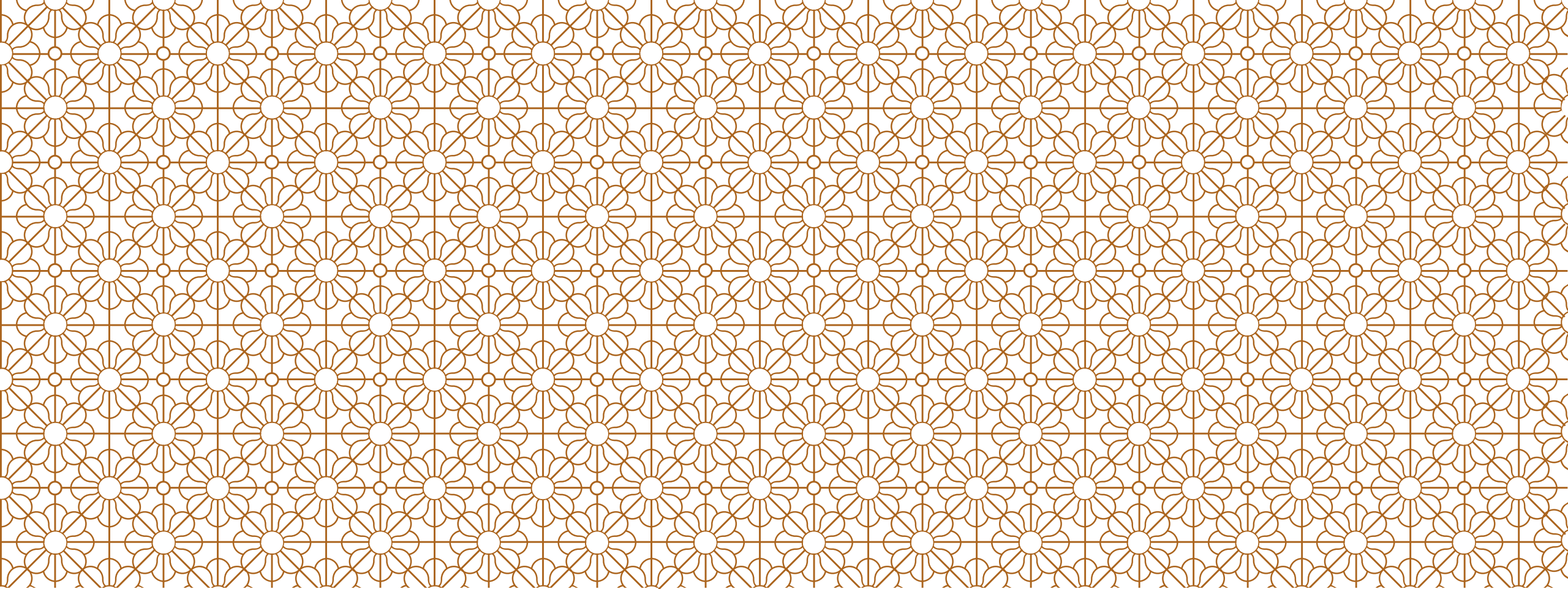
という定義は一般には受け入れられているが・・・

- その脱文法範疇的な性格から、形態統語論的特徴は多様。
- その用法や一つ一つの機能の分類に関して、見解の一致に至っていない。
- 談話標識のリストが限定できるかどうかが議論の的。
- そもそも、談話をめぐる現象は複雑。コーパスデータの性質により分布は可変的。

研究設問

以下が本研究のリサーチクエストである。

- RQ1 談話標識はどのような基準や条件によって認定が可能か？
- RQ2 談話標識は閉じられたリストなのか？
- RQ3 談話標識は一つの機能的カテゴリーとして認定可能か？



調査手法



分析データ

TUFS 2010のうち、CÉFC に収録されているコーパスデータ 5 つを選択

コーパスID	延べ語数
会話ID: 01	13226
会話ID: 03	16961
会話ID: 04	16019
会話ID: 05	18303
会話ID: 09	17220

調査の手順

データの入手

CÉFC からテキストを入手



コーパス駆動型調査

各ファイルの最初の500例を網羅的に抽出⇒計2500例の表現を抽出



分析

統語的・機能的観点から各表現の特徴を分析

CRIBLE (2018) の定義

■Crible (2018: 62) の談話標識の定義

- 文法カテゴリーが多様 (grammatically heterogeneous)
- 統語的に任意 (syntactically optional)
- 多機能的 (polyfunctional)
- 談話解釈を方向づける 手続き的な手がかり (procedural cues)
- 4つの機能領域の一つを担う
- 統語的・意味的自律性をもつ (syntactic and semantic autonomy)

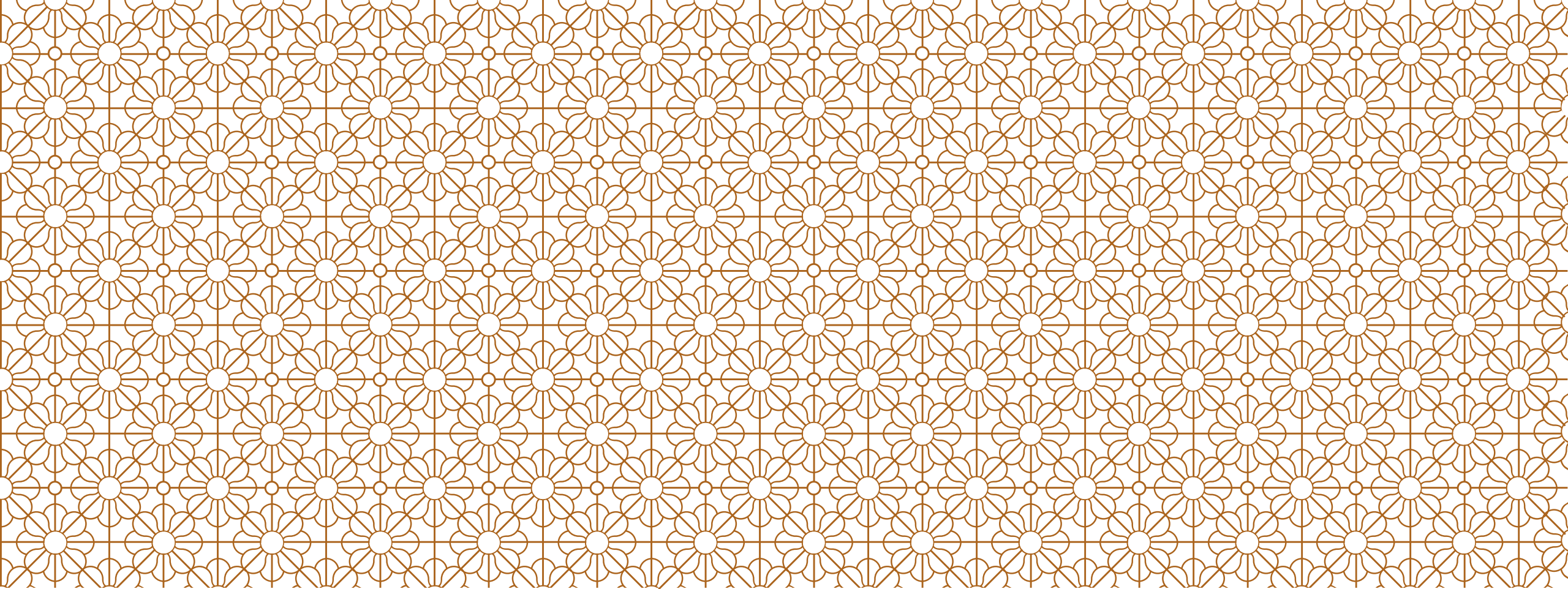
→ この定義に基づき、Crible (2018) は、談話標識の網羅的なリストを作成。

抽出対象の表現

- 本調査では、以下の要素を網羅的に抽出
 - ・ Crible (2018) のリスト掲載表現
 - ・ リスト未掲載の表現
 - Crible (2018) の定義に基づきリストに掲載されうるリスト未掲載表現
 - 解説的語用論標識 (Fraser 1996a, 1996b, 小野寺 2024)
→ franchement, je crois, je pense など
 - ・ パラ言語的要素 (euh, ah, oh, mh など)

アノテーション内容

カテゴリー		アノテーション内容
タイプ	抽出表現の分類	パラ言語 / リスト掲載 / リスト未掲載
削除可能性	削除しても支障がないか	可能 / 不可能
位置の移動性	他の位置に移動できるか	可能 / 不可能
機能領域	Crible (2018) の機能分類	観念的領域 (ID) / 修辭的領域 (RH) / 連鎖的領域 (SE) / 対人関係調整領域 (IN) / 該当無し
共起関係	他の談話標識との共起	有 / 無



分析結果



頻度調査（タイプ頻度）

コーパスID	タイプ頻度
会話ID: 01	89
会話ID: 03	93
会話ID: 04	56
会話ID: 05	76
会話ID: 09	63
全体	174

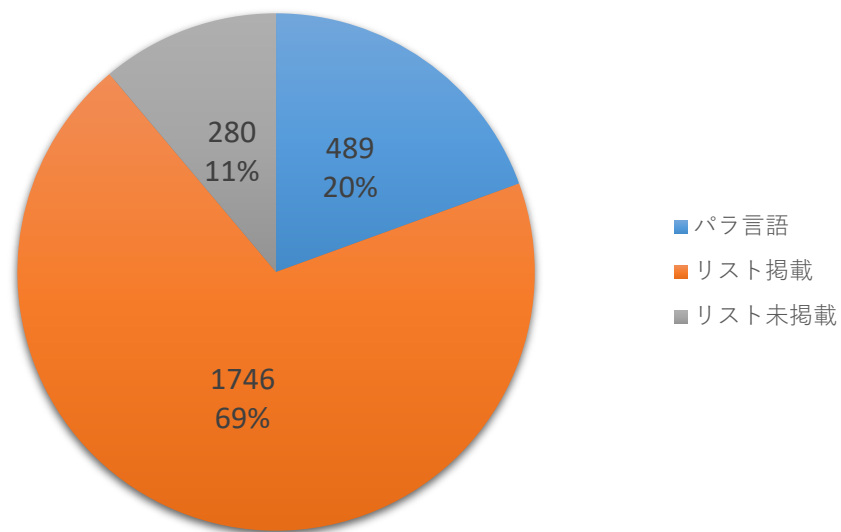
抽出された表現

- ・ 頻度の高い表現は以下の通りである

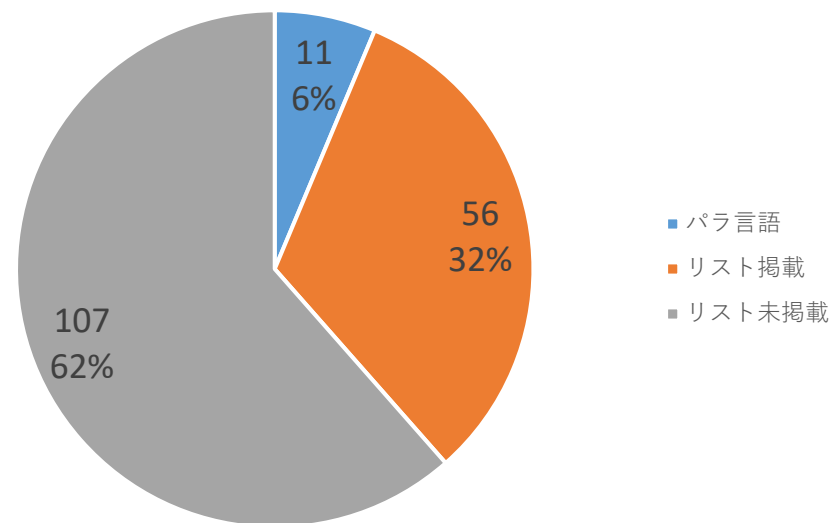
順位	MD	頻度	順位	MD	頻度
1	euh	269	16	voilà	42
2	ouais	246	17	si	36
3	mais	188	18	en fait	33
4	oui	171	19	d'accord	29
5	et	153	20	alors	28
6	ah	116	21	quand	28
7	non	99	22	enfin	25
8	ben	79	23	puis	24
9	bon	74	24	c'est vrai que	23
10	parce que	73	25	quand même	23
11	donc	59	26	tu vois	17
12	hein	53	27	eh	16
13	mh	50	28	comme	14
14	quoi	50	29	oh	14
15	après	43	30	c'est ça	12

CRIBLE (2018)の談話標識のリストとの比較

全体のトークンにおける割合



タイプ頻度における割合



総頻度においては、リスト掲載表現が7割近くを占める。
タイプ頻度では、リスト未掲載表現が多数を6割を占める。

リスト未掲載表現

c'est vrai que	23
c'est ça	12
je sais pas	12
je crois	11
c'est vrai	11
c'est sûr	9
de toute façon	9
finalement	8
OK	8
effectivement	6
aussi	5
je crois que	5
je pense	5
normalement	5
waouh	5
ensuite	4
évidemment	4
surtout	4
surtout que	4
au fait	3
bien sûr	3
c'est clair	3

- Crible (2018) の談話標識の未掲載表現

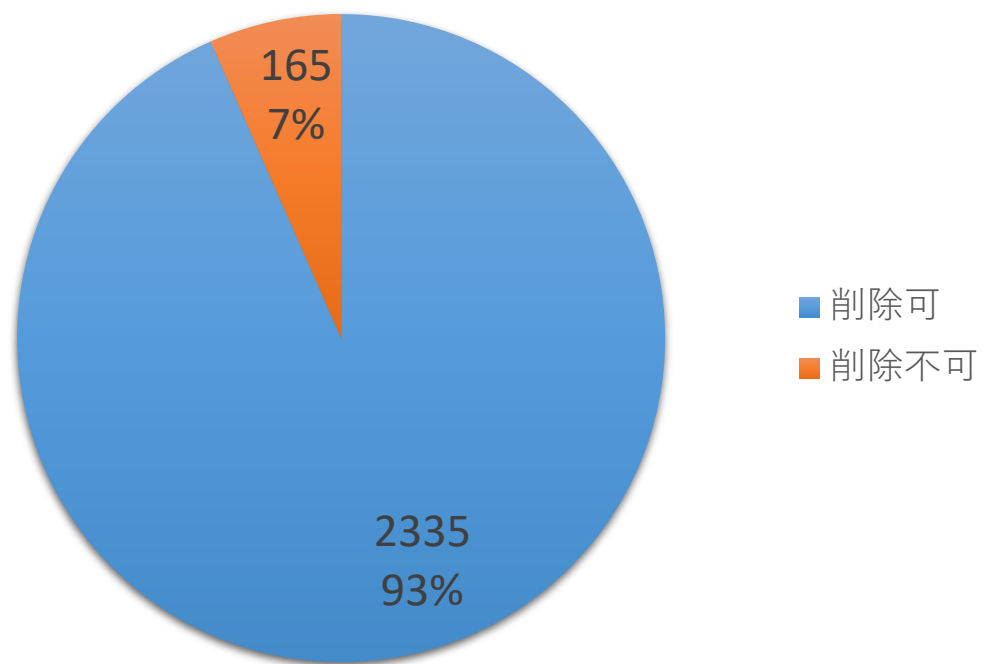
→107例

- 一度のみ出現の Hapax legomenon

→56例

oui mais je veux dire oui mais **ce que je veux dire c' est que** au fur et à mesure ça ça me fait bien cinq cents euros par mois là (ID: 01)

削除可能性



- 談話標識の特徴としての統語的随意性 (syntactic optionality)
⇒実際には、削除が困難、あるいは削除できない表現も存在する

- Crible (2018) のリストにある quand ⇒ 従属節を導く接続詞

NH: ça rajoute eh il faut les harceler hein

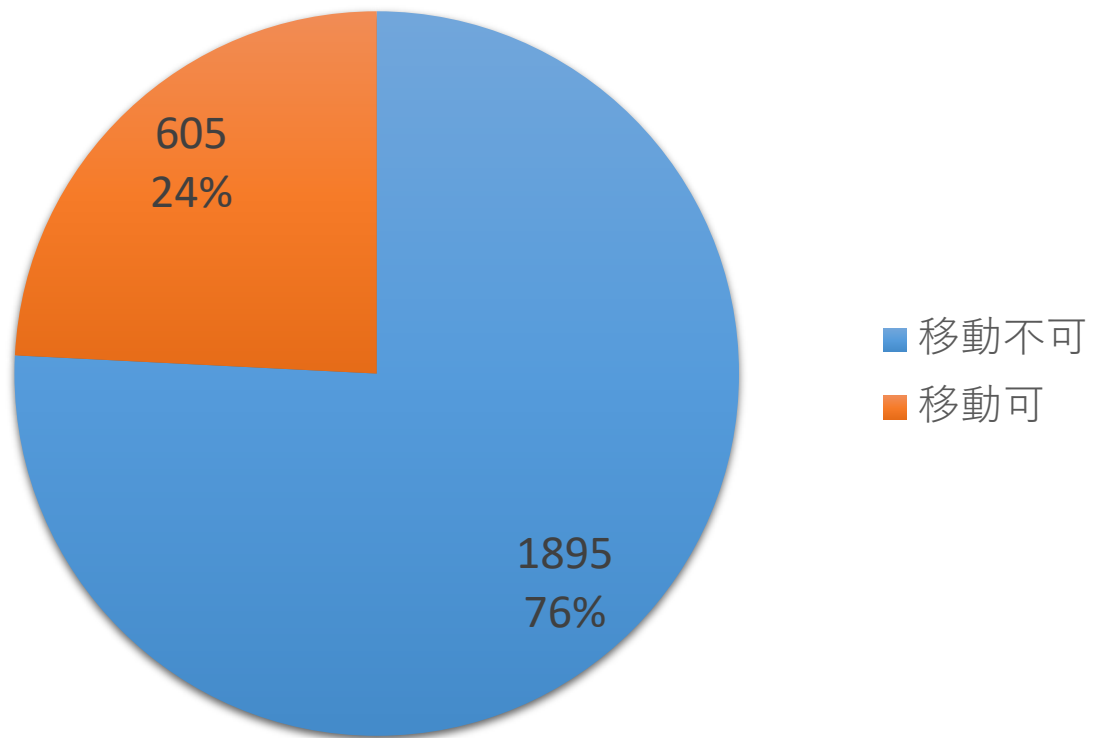
OG: quand ça marche pas (ID: 01)

- mais は、談話構成上重要な役割を果たす。⇒削除困難な場合

NH: essaye de voir si il y a pas des des fois il y a des studios vachement sympas

OG: oui **mais** ça coûte trop cher ici tu vois (ID: 01)

位置の移動性



移動可能な談話標識

- 談話標識は、発話の様々な位置に現れる (Dostie & Pusch, 2007: 4)

⇒実際には、様々な位置に移動できる標識は限定的

- 移動可能な表現：意味や機能的変化はあるが、様々な位置に出現可能

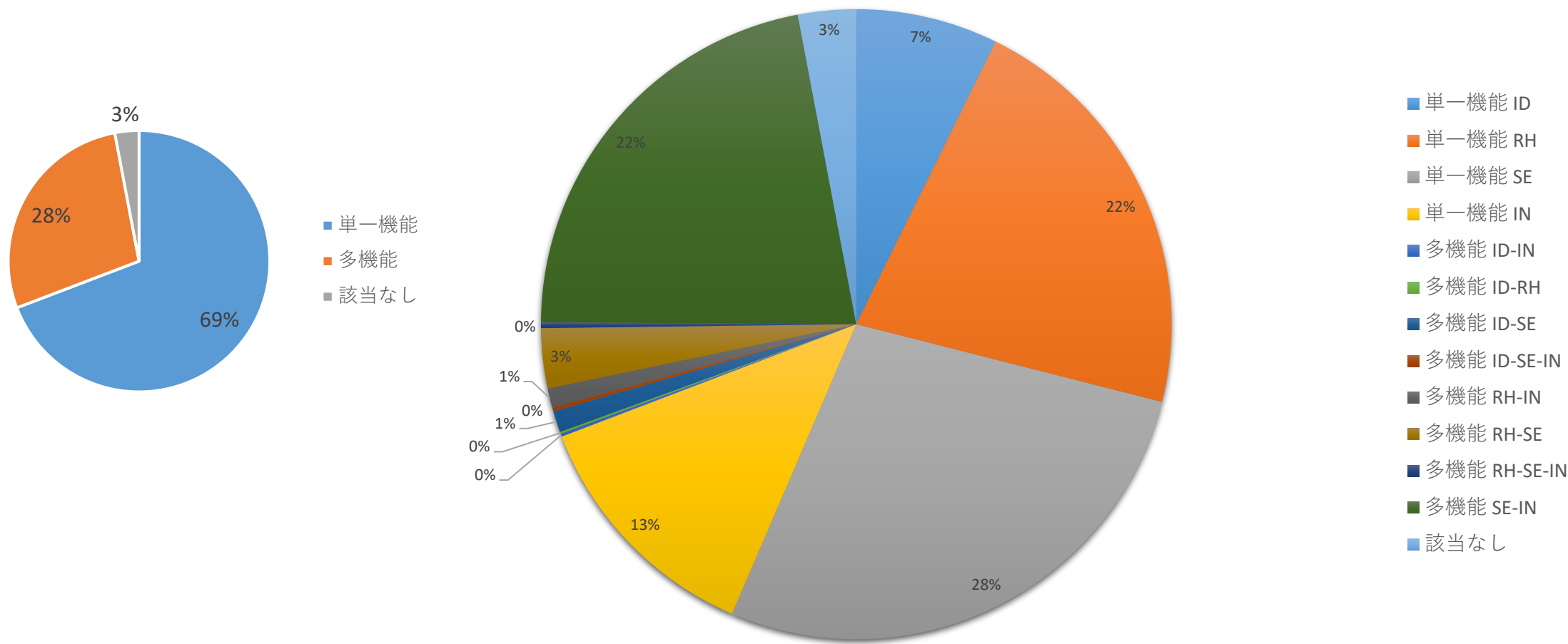
c' est un génie **quand même** mais au passage **euh** s' il peut **euh** dire quelques remarques qui font pas toujours plaisir **d' ailleurs** hein (ID:05)

- 移動不可能な表現：mais は発話頭、hein は発話末から移動不可。

c' est un génie quand même **mais** au passage euh s' il peut euh dire quelques remarques qui font pas toujours plaisir d' ailleurs **hein** (ID:05)

機能分類の結果

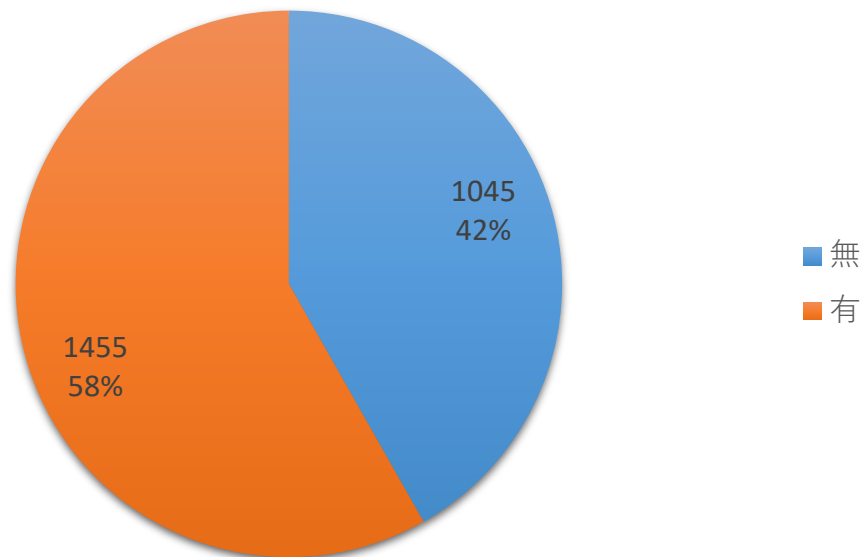
パラ言語 484 例は本分析からは除外している。



機能分類の結果

- 単一機能の用例が6割以上
 - そのうち、連鎖的領域 (SE) は28 %、修辭的領域 (RH) は22%
 - リスト未掲載の表現は、対人関係調整領域の機能を担う表現が多い
- 多機能的な談話標識は28%を占める
 - 最多は連鎖的領域 (SE)-対人関係調整領域 (IN) の組み合わせ

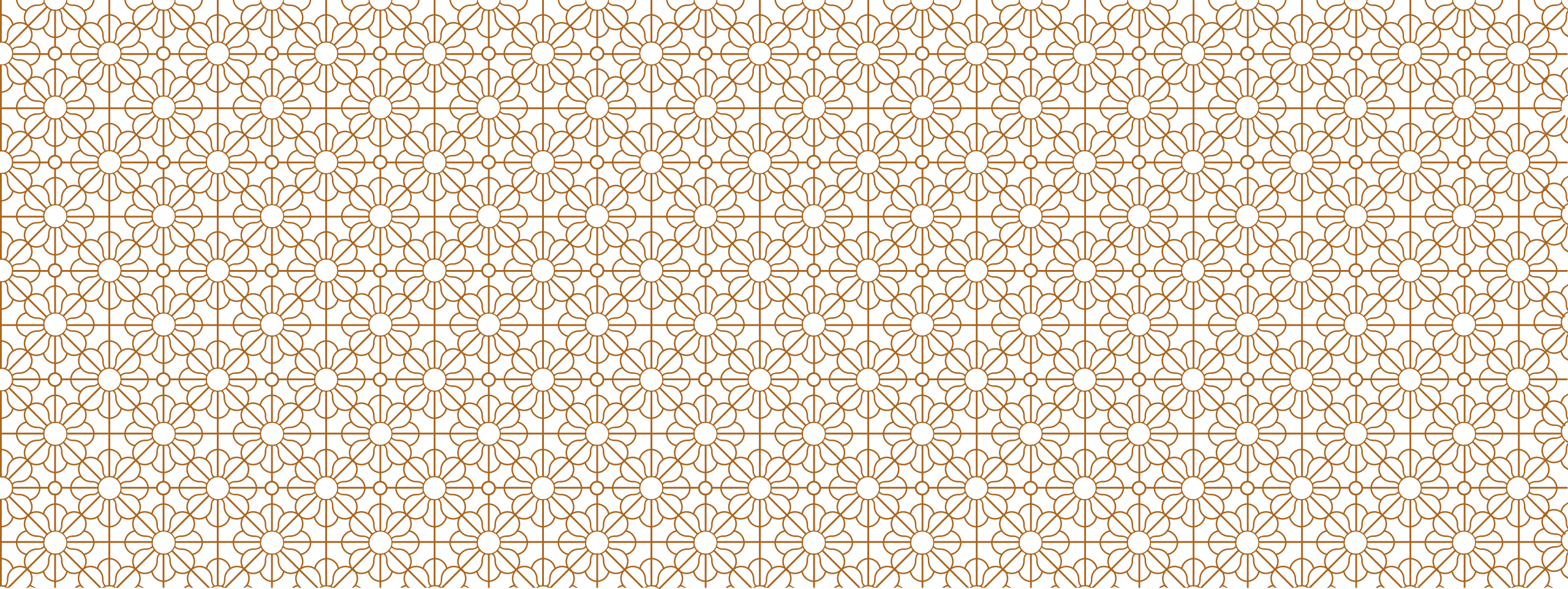
共起の有無



半数以上の例が他の談話標識と共起している

共起しているものの例

ah ouais
ouais ouais
donc euh
ah oui
mais bon
et puis
oui mais
non mais
ouais mais
euh oui
mais euh
ah ben
ah mais
ah oui oui oui
eh ouais
ah ben oui
ah et voilà
ah oui oui
et euh
et voilà
euh bon
euh ouais
hé ben
non non non
ouais ben
ouais ouais ouais
oui ben oui
quoi donc euh



考察と結論

研究設問への答え

RQ1 談話標識はどのような基準や条件によって認定が可能か？

- 音韻論的基準：（今回は調査対象としていない）
- 統語論的基準：
 - ✓ 脱文法的範疇（＋）
 - ✓ マクロ統語論的依存関係（＋）
 - ✓ 凝結性（＋）複数の語彙素から成る談話標識の場合は凝結性（語彙化）が見られる。
 - ✓ 移動可能性（＋／－）マクロ統語論レベルでは位置が変わると機能が代わる可能性がある。
 - ✓ 削除可能性（＋／－）削除前後の発話が等価とは限らない。
 - ✓ 共起可能性（＋／－）共起関係にある要素が固定されたものであるか検討の余地あり。

意味・語用論的機能：

- ✓ 手続き的意味（＋）
- その他
 - ✓ 頻度（＋）

RQ2 談話標識は閉じられたリストなのか？

□高頻度で現れるパラ言語学的単位（*eah, mh*など）の扱いは難しい。

✓これにジェスチャーの問題も加わるとさらに複雑

□語彙的表現のリストを限定することは難しい。

✓コンテキスト次第では手続き的意味を持ち得る表現が多数存在する。

✓*Hapax legomenon* のケースが数多く見られる。（コーパス内に一度しか現れなかった要素。談話標識として認定するか、談話標識化した用法とみるか？コーパス規模と質の限界も考慮。）

□社会言語学的変異が大きく、流行り、廃れも多いことが予想される。リストの固定化は難しい。

RQ3 談話標識は一つの機能的カテゴリーとして認定可能か？

- 談話構築に寄与する、という大きなカテゴリーとしては、認定可能。
- 統語的特徴、意味機能による分類も可能。
- しかし、1つの談話標識が一つの分類に収まらないことが多い。

→多くの談話標識は多機能的。

結論

- 今回は、Crible (2018) に基づき、4つの領域で機能を分類
→ より詳細な談話標識の機能的分析が必要
- 本発表は、2500例についての一次調査の結果の報告
→ さらなる例を抽出し、分析する必要

今後の課題

- パラ言語学的単位、音韻論的特徴を含め、今後考える必要あり。学習者の母語転用が現れやすいのではないか。
- 教育的観点で見た場合、インフォーマルな学生同士の会話だけではなく、改まった場での予備原稿のあるプレゼン、SNSの投稿、メールの書き方など、様々な状況に応じた談話標識の使用を学習者は学んでいく必要があるのではないかとと思われる。様々なジャンルごとでの対照研究が必要。

謝辞

本研究はJSPS科研費 JP25K00472、JP25K23001の助成を受けたものです。

参考文献

- Anscombre, J-C. (2024) Les marqueurs de discours. Une catégorie linguistique ? *Scolia* 38, <https://journals.openedition.org/scolia/2893>
- Beeching (2015) Variability in Native and Non-Native Use of Pragmatic Markers: the Example of *Well* in Role-Play Data, *Researching Sociopragmatic Variability*, Springer, 174-197.
- Berrendonner, A. (1990) « Pour une macro-syntaxe », *Travaux de Linguistique*, 21, 25-36.
- Blanche-Benveniste, C. et al. (1990) *Le français parlé, études grammaticales*, Paris, Éditions du CNRS
- Broisson & Degand (2022) How egocentric is discourse marker use? Investigating the impact of speaker orientation and cognitive load on discourse marker production, (in) Cuenca & Degand (2022) *Discourse markers in Interaction*, De Gruyter, 121-158.
- Crible, L. (2018) *Discourse Markers and (Dis) fluency: Forms and functions across languages and registers*. Amsterdam / Philadelphia, John Benjamins.
- Crible, L & Degand, L. (2019) Domains and Functions : A Two-Dimensional Account of Discourse Markers. *Discours*, 24. 3-35.
- Degand, L. & Evers-Vermeul, J. (2015) Grammaticalization or Pragmaticalization of discourse markers? More than a terminological issue, *Journal of Historical Pragmatics*, 16 (1), 59-85.
- Dostie, G., Pusch, C. (2007). Présentation. Les marqueurs discursifs. Sens et variation. *Langue française*, 2(2), 3-12. <https://doi.org/10.3917/lf.154.0003>
- Ducrot, O. (1998) *Les mots du discours*, Paris : Editions de Minuit.

- Fraser, B. (1996a). Pragmatic markers. *Pragmatics* 6(2)
- Fraser, B. (1996b) Commentary Pragmatic Markers in English, Boston University
- Fraser, B. (1999) What are discourse markers? *Journal of Pragmatics* 31, issue 3-4, 931-952
- Hansen, M, B-J (1998) The semantic status of discourse markers, *Lingua* 104, 3-4, 235-260.
- Heine, B., Kaltenböck, G., Kuteva, T., Long, H. (2021) *The Rise of Discourse Markers*, Cambridge University Press.
- Kerbrat-Orrechioni, C. (2005) *Le discours en interaction*, Collection U, Lettres, Linguistique. Paris: Armand Colin.
- Paillard, D. (2017) Scène énonciative et types de marqueurs discursifs, *Langages* 207, 17-32.
- Pons Borderia (2008) Do discourse markers exist? On the treatment of discourse markers in Relevance Theory, *Journal of Pragmatics* 40, issue 8, 1411-1434.
- Redeker, G. (1990) Ideational and pragmatic markers of discourse structure, *Journal of Pragmatics* 14, issue 3, 367-381.

Shiffrin, D. (1987) *Discourse markers*. Cambridge University Press.

Traugott, E.C. (2022) *Discourse structuring markers in English: A historical constructionalist perspective on pragmatics (Constructional Approaches to Language 33)*. Amsterdam / Philadelphia: Benjamins.

小野寺典子 (2024) 『談話標識へのアプローチ 研究分野・方法論・分析例』, ひつじ書房.
コーパス

ORFÉO (Outils et Ressources sur le Français Écrit et Oral)
<<https://repository.ortolang.fr/api/content/cefc-orfeo/10/documentation/site-orfeo/home/index.html>> (最終アクセス2026/6/13)

古賀健太郎, 秋廣尚恵, 川口裕司 (2012) 「Aix 話し言葉コーパスプロジェクト」
『ふらんぼー』 37, 東京外国語大学フランス語研究室, 37-54.