

ゆずこ 電車、久しぶりに乗るね～
唯 あー、そうかもなー
ゆずこ いつぶりくらいかな
縁 去年の海？
唯 いやいや、さすがにもっと乗ってるだろ
ゆずこ 電車、あんまり乗らんよね～
縁 あんまりね～

ゆずこ いろんな人がいるね
縁 うん
ゆずこ なんか、さ
縁 うん？
ゆずこ いろんな人がいるんだなーって。思うよね？
縁 あー、わかるー
唯 …。
唯 あんまじろじろ見るなよ
縁 えー？
唯 ゆずこ、他の人のことじろじろ見てただろ？
ゆずこ なにそれ？
縁 ？
ゆずこ わたしだけを見てろってこと？
唯 いやそうじゃなくて…
縁 あ、席空いたよー
ゆずこ 座る？
唯 …。

縁 かわいかったねー
ゆずこ かわいかった！
唯 そうだな
縁 でもすごいぐずってたねー
ゆずこ ぐずってた！
唯 （なんか、ゆずこみたいだった…）
ゆずこ …？

唯 なんだよ？
 縁 なんかねー
 ゆずこ ん？
 縁 今の赤ちゃん、ゆずちゃんみたいだった
 ゆずこ ほへ？
 唯 あー、うん...
 ゆずこ えー？ なにそれー？ それ喜んでいいの？
 縁 なんかー、しゃべり方とか？
 唯 喋ってたか？
 縁 だーだーって
 ゆずこ それ、バカにされてる気がするんですが...

ゆずこ 喃語。乳児の発する言葉。言語を獲得する前段階
 唯 さっきの赤ちゃんだ
 縁 わんわんとか？
 ゆずこ んーと。幼児語。乳幼児の会話に用いられる言葉
 縁 ふーん？
 ゆずこ わんわんとかは幼児語っていうんだって
 ゆずこ 言語の獲得...
 唯 あ、これテレビで見たことある
 唯 ブーバ・キキ効果
 唯 2つの図形を見せて、
 唯 どっちがブーバで、どっちがキキかを答えさせるという心理実験
 縁 ブーバー？
 唯 うーんと、言葉に意味はなくて、音の響きとかから判断するんだって
 ゆずこ ぶ～ば～
 縁 なんかー、キキの方が痛そうな感じる
 ゆずこ 黒板みたいな？
 縁 あー、確かに
 唯 ん？ どっちが黒板？

ゆずこ E. Mark Gold さん ^[1]
 縁 誰？
 ゆずこ 幼児の言語獲得の定式化を試みた人、だってー
 ゆずこ 幼児は親が話す言葉だけから言語を学習する...

唯 そうだな？
 ゆずこ 正しい言葉だけから正しい言語を学習できる
 ゆずこ これを Learning from positive data 、と言うそうです
 唯 それっ当たり前じゃない？ 何が難しいんだ？
 縁 うーん、よくわからない
 ゆずこ 子供は学んだ言葉を使って新しく作文するかもしれない
 ゆずこ それが正しい言葉であるかどうかを、親の反応から学習する
 ゆずこ この場合は、「正しくない言葉」も学習できるチャンスがある
 縁 ふーん？
 ゆずこ あ、こんな例があるよ¹

Q. 次の A は自然数 $(1, 2, 3, \dots)$ からなる集合です. A はどんな集合ですか？

$$A = \{2, 4, 6, 8, 10, \dots\}$$

唯 ふーん？
 唯 いやいやいや、偶数だろ？
 縁 うんうん

A. A は偶数全体.

$$A = \{2, 4, 6, 8, 10, 12, 14, \dots\}$$

ゆずこ ざんね～ん
 ゆずこ とは限りません

A. A は偶数または 13 の倍数からなる集合.

$$A = \{2, 4, 6, 8, 10, 12, \mathbf{13}, 14, \dots\}$$

縁 えー、インチキだよー

¹参考文献 [3] の例を改変

ゆずこ 最後まで聞かないほうが悪いのです
 唯 お？ 急になんだ？
 ゆずこ そうじゃなくってー
 縁 あー、「...」ね
 ゆずこ そう！「...」なのです
 唯 ふーん？「聞く」ってなんだ？
 ゆずこ ゆかりちゃん、やってみて
 縁 ええっと
 縁 「12 は A ですか？」
 ゆずこ はい。そーです。
 縁 「14 は A ですか？」
 ゆずこ はいっ。そーですっ
 唯 それだとやっぱり偶数じゃないか。2 の倍数以外も聞かないと
 唯 13 は？ って聞いて、、、
 ゆずこ 13 も A だ、って答えるよ
 唯 でもそれでも、有限の質問だけから決めるのは無理じゃないか??
 唯 全部の数について聞かないと、、、
 ゆずこ 「これは A に含まれる」っていうのを “informant” っていうらしいよ
 縁 「含まれない」ってのは？
 ゆずこ うーん。「含まれる」を positive data とも呼ぶ、とも書いてある...

- (有限とは限らない) 集合で、学習したい対象のことを Concept という。例えば一つの言語 (日本語とか英語とか) は Concept の一例。例えば自然数の部分集合は Concept の一例。
- ある要素が学習したい Concept に属するという情報を informant (正データ, positive data とも) と呼ぶ。例えば親が話す言葉は informant. 例えば自然数の部分集合 A について $x \in A$ は一つの informant.

唯 ってことは、さっきの自然数の集合が言語にあたるのか...？
 縁 自然数が言葉ー？

ゆずこ そこで、「極限同定」という発想が生まれるわけです

唯 うん？

次のような Concept (言語) の学習の枠組みを「極限同定」と呼びます。

- ある Concept をこれから学ぼうとする学習者がいます。
- 学習者は一つの informant を得ます。「これは学習すべき Concept に含まれる」という情報であって、「含まれない」という情報は得られないことに注意してください。
- 学習者はそれまでに得た informant を用いて、1 つ Concept を推論します。それが学習者の学習結果です。
- 以上を 1 つのステップとして、これを何度も繰り返します。^a

更に次のようなとき、極限同定が「成功した」と言います。

- あるステップで真の Concept を推論し、かつ、
- 以降のステップでは、常にその Concept を推論する。

^aこの情報の得方を正提示 (positive presentation) と言いますが、厳密にはもっと強い制約があります。それは Concept の任意の要素はいつかは必ず informant として提示される、というものです。

縁 じゃあさっきの例で言うと、

縁 2, 4, 6, 8, 10, ... ってのが正提示でー

唯 うんうん

縁 数字一つを受け取るごとに、自然数の部分集合を一つ答える

唯 推論な

ゆずこ ちなみに、推論した Concept の列を「推論列 (guessing sequence)」っていうよ

正提示 2, 4, 6, 8, 10, ... に対して、推論列 $A_1, A_2, A_3, A_4, A_5, \dots$

2,	4,	6,	8,	10,	...
↓	↓	↓	↓	↓	
$A_1,$	$A_2,$	$A_3,$	$A_4,$	$A_5,$...

ゆずこ あ、ちなみにこの正提示は、小さい順に並べてるけど、

ゆずこ 別にそれに意味はなくて、順序はどうでもいいし、それに重複してもいいんだって

縁 2, 10, 2, 10, 8, ... みたいに？

ゆずこ そうっす

唯 時刻 t の推論を A_t って書いたんだな

唯 じゃあさっきの「極限同定の成功」の条件は

$$\lim_{t \rightarrow \infty} A_t = A$$

唯 って書いていいのか？

ゆずこ うーん、いいんじゃないかな

ゆずこ 文字通り形式化すると $\exists T, \forall t > T, A_t = A$ だけど

ゆずこ 極限同定が成功したとき、Concept を学習したって見做すらしいよ

縁 それで？

ゆずこ うん？

縁 これで、人間の言葉が学習できるの？

ゆずこ うーん、、どうなんだろう？

縁 えー

唯 どうなんすか？ ゆずこさーん

ゆずこ むーん。。。.

ゆずこ あ！

唯 なになに

ゆずこ ”文脈自由文法の学習は不可能である”^{[1] 2}

縁 えーなんでー？

唯 推論の具体的な方法も決めてないのに

ゆずこ TABLE 1 に書いてるのだとー...

唯 なになに？

ゆずこ 文脈自由文法も文脈依存文法も正規言語も、正提示からの極限同定は不可能

縁 だからなんでー？

ゆずこ うーん、詳しい記述はどこにあるんだろう

²英語は文脈自由文法であると言われる。いや、少なくとも反証はされていない。

ゆずこ あ、あったあった

ゆずこ "super-finite class of languages"

縁 クラス？

ゆずこ うん、言語族、のことかな？

唯 そんなの今まで話に出てこなかったぞ

ゆずこ ごめんごめん。説明がめんどそうだったから...

ゆずこ 「自然数の部分集合」とか「自然言語」みたいに、推論する Concept を中から選べるための枠組みみたい

縁 え？ じゃあ、推論って、選択のこと？

唯 なんか急に問題が簡単に見えてきたなあ

列挙による推論

Concept のクラスが添字つき集合^aの場合、列挙による推論がありえる。この方法ではまず Concept を列に並べる。

$$\mathcal{C} = \{C_1, C_2, C_3, \dots\}$$

次のような方法で推論を行う。

- 添字 $i = 1$ を持っておく。
- 一つの informant を受け取る。
- 過去に受け取った informant のどれかと C_i とが矛盾するか調べる。
- どれかと矛盾するなら、 C_i は正しい Concept ではないので、添字 i を 1 増やす。矛盾しないなら何もしない。
- 推論として C_i を答える。

^a集合の要素をもれなく列として並べられるということ

縁 さっきの自然数のは？ これで解ける？

唯 いやいや、無理だろ

縁 どうして？

唯 自然数の部分集合全体は可算じゃないから

縁 そっかー

唯 この方法で推論するには「あり得る言語が多すぎる」ってわけだな

ゆずこ もっと単純なやつじゃないと、これはダメみたいだね

縁 なにならいいんだろう
 唯 「ある数の倍数」とかどうだ？
 縁 それなら可算だね

倍数全体という Concept クラス

自然数 n の倍数全体を \mathbb{Z}_n と書く. $\mathcal{C} = \{\mathbb{Z}_n : n \in \mathbb{N}\}$ は列挙による方法で
 正提示から極限同定可能?

ゆずこ うーん...
 唯 いや、無理だな
 縁 えー、なんでー？
 唯 だってほら、答えを「4の倍数」だとするだろ
 唯 正提示としては 4, 8, 16 って感じで、4の倍数だけが来るけど、
 唯 4の倍数って2の倍数でもあるだろ？
 縁 つまり？
 ゆずこ ほら、2の倍数の可能性を捨てきれないわけじゃん
 唯 単純に、 $\mathcal{C} = \{\mathbb{Z}_1, \mathbb{Z}_2, \dots\}$ って並べてたら、 $i = 2$ で停まっちゃう
 ゆずこ 添字の順序を工夫すれば... うーん...
 唯 \mathbb{Z}_4 を \mathbb{Z}_2 の前に持ってくればいいんだけど、そしたらキリがないし
 縁 あーでも、ちょっとズルをしたら上手くいくかも
 唯 どんな？
 縁 あのねー、正提示の一番最初だけ先に見ちゃうの
 縁 で、それが m だったら、

$$\mathbb{Z}_m, \mathbb{Z}_{m-1}, \dots, \mathbb{Z}_2, \mathbb{Z}_1$$

縁 って並べるの
 唯 あー、それなら上手くいくなー
 唯 上限を見つけて、そこから降りてくわけだな
 ゆずこ あ、ていうか！
 ゆずこ これって、提示される数の最小値を取ればいいだけだね
 縁 あー、確かに
 唯 提示される数は整数だから、絶対値は取らないとだけだな

自然数 n の倍数全体を \mathbb{Z}_n と書く. Concept class, $\mathcal{C} = \{\mathbb{Z}_n : n \in \mathbb{N}\}$ は次のような方法によって正提示から極限同定可能.^a

- 数 $m = \infty$ を持つ
- informant m_t を受け取る
- m より $|m_t|$ の方が小さい時、 m を $|m_t|$ で上書きする
- Z_m を推論

^aこれは正しく、極小言語 (minimal language; MINL) 戦略の自然数バージョン

縁 んー。それはいいんだけどー
 縁 赤ちゃんは？ 言語はー？
 唯 あ、ていうか、文脈自由文法は？ ”super-finite” は？
 ゆずこ あ、そうだったそうだった。忘れてました

唯 正提示から学習できる言語。パターン言語 [2]
 ゆずこ お？
 唯 んんん?? $\Sigma A \dots$?
 ゆずこ どれどれ
 唯 んー。パターン言語の定義らしいんだけど...
 ゆずこ ああ、たぶんこれ、Kleene 閉包だよ
 唯 こんなヘンな記法が...

非形式的に「パターン」を説明します。日常会話で用いるいわゆる「パターン」とは違う意味で使うので注意して下さい。ここで言うパターンとは、要は、「空欄のある文」です。

- 例. “世界が でありますように”

また空欄には「名前」をつけることができます。

- 例. “ (x) が (x) を (y) でした”

非形式的に「パターン言語」を説明します。パターン言語とはある1つのパターンによって「説明される」言語です。

先ほど例に挙げたパターン“ (x) が (x) を (y) でした”を再び用います。このパターンが「説明する」言語とは、空欄を自由に埋めて出来る文からなる集合のことです。ただし、同じ名前の空欄には同じもので埋めなければいけません。

- “ 私 が 私 を 見つめ ました”
- “ 深淵 が 深淵 を 覗いて ました”

縁 急に日本語だー
 唯 同じ名前の空欄があるって、同じものが何度も出現するってことだろ？
 唯 あんまり自然言語にそういうのって出てこないような
 縁 同じ名前の空欄が出てこないものは「正則パターン言語」
 ゆずこ あ、これってもしかして、深さが1しかない文脈自由文法？
 唯 あー、なるほど
 唯 ん？
 ゆずこ あ！ ほら見て！ 鎌倉！
 縁 大仏はー？
 ゆずこ うーんどこだろう

唯 こっからじゃ見えないんじゃないか？
縁 でもじゃあ、もうすぐ！ だね
ゆずこ うん！

参考文献

本物語は次の参考文献をヒントに創作しました.

- [1] E. Mark Gold: “Language Identification in the Limit”, in *Information and Control* 10 (1967)
- [2] Angluin: “Positive Inference of Formal Languages from Positive Data”, in *Information and Control* 45 (1980)
- [3] Hiroki Arimura, Takeshi Shinohara and Setsuko Otsuki: “Finding Minimal Generalizations for Unions of Pattern Languages and Its Application to Inductive Inference from Positive Data”, in *In Proc. the 11th STACS, LNCS* 775 (1994)
- [4] @cympfh: “言語の極限同定みたいな話”,
<http://cympfh.cc/study/language-identification/history/it.pdf>