

言語学とデータサイエンスに関する
ワークショップ・
参加無料

言語データとその「鏡」： 機械学習モデルを用いた 言い誤りと失語症例の 分析

2023年
12月23日〔土〕
13:30 ～ 17:00

①認知モデルの説明1：健常者	寺尾 康（静岡県立大学）
②認知モデルの説明2：失語症者	上間清司（武蔵野大学） 橋本幸成（目白大学）
③機械学習モデルの説明	浅川伸一（東京女子大学）

認知モデルの紹介

心理言語学において提案されてきた代表的言語産出モデル

- 70年代 Garrettの二段階モデル
- 80年代 相互活性化モデル
- 90年代 Leveltモデル

Garrettの二段階モデル 1

当時、MIT教授だった M.Garrettは、自ら手書き収集した
日常発話に生じた言い誤り6000例をもとに、“文産出モデル” を提案
(Garrett 1975)

3種類の交換型言い誤り

- (i) a. I left my briefcase in my cigar.
b. ...which was parallel to a certain sense in an experience...
- (ii) a. I went to get a cash checked. (check cashed)
b. It waits to pay. (pays to wait)
- (iii) a. We expected Jom and Terry to be here. (Tom and Jerry)
b. It's past fassing by... (fast passing)

Garrettの二段階モデル 2

	同一品詞	内容語	文法語	同一句	離れた二語で
(i)タイプ 語交換	yes	yes	yes	no	Yes
(ii),(iii)タイプ 形態素、音韻交換	no	yes	no	yes	no

(i)タイプは、順序の制約よりも文法的な制約が強い。

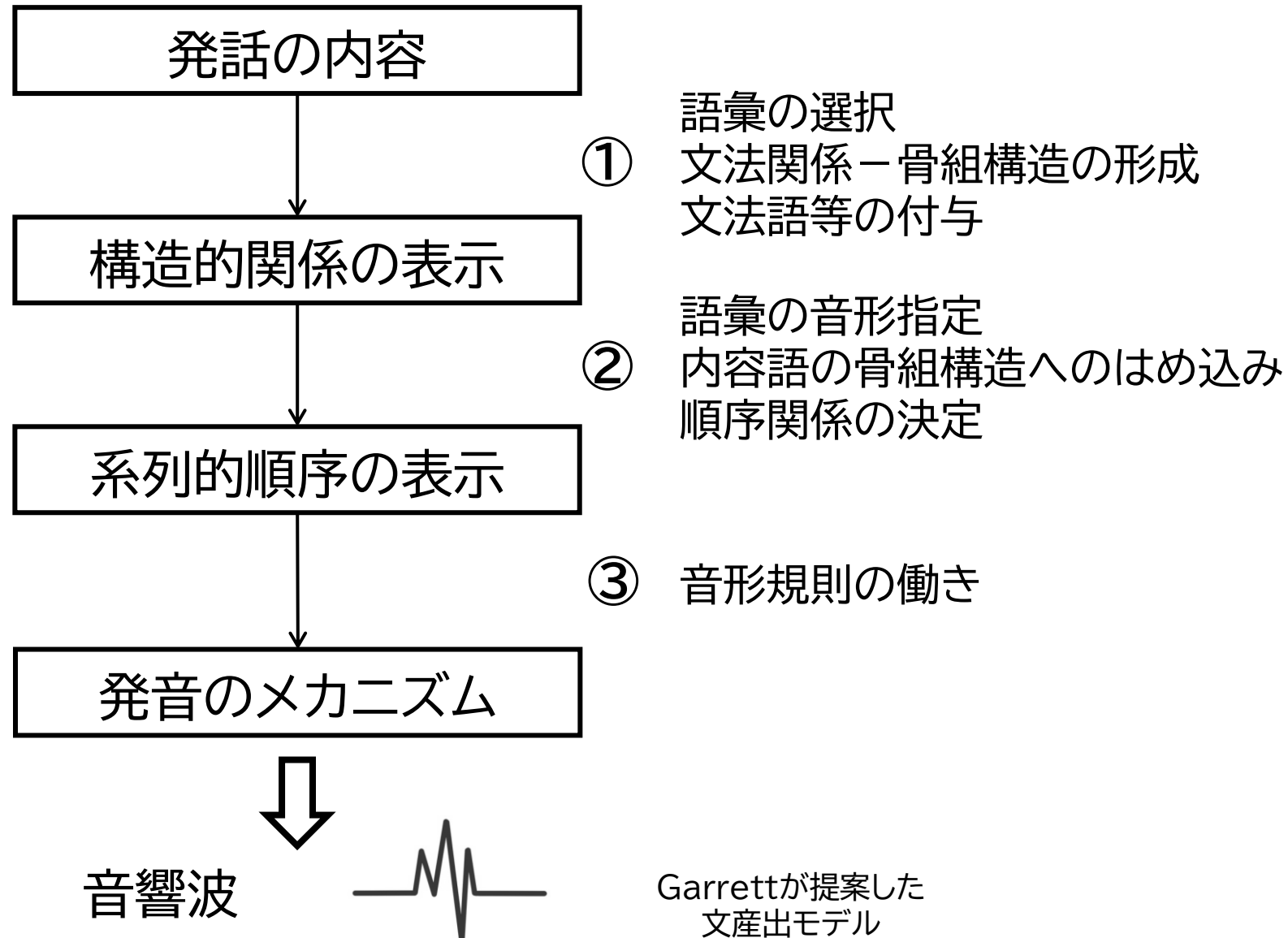
(ii),(iii) タイプは、その逆。

系列的な位置が決められるレベルで起こっている可能性が高い。

そこで、彼が提案したのは…

Garrettの二段階モデル 3

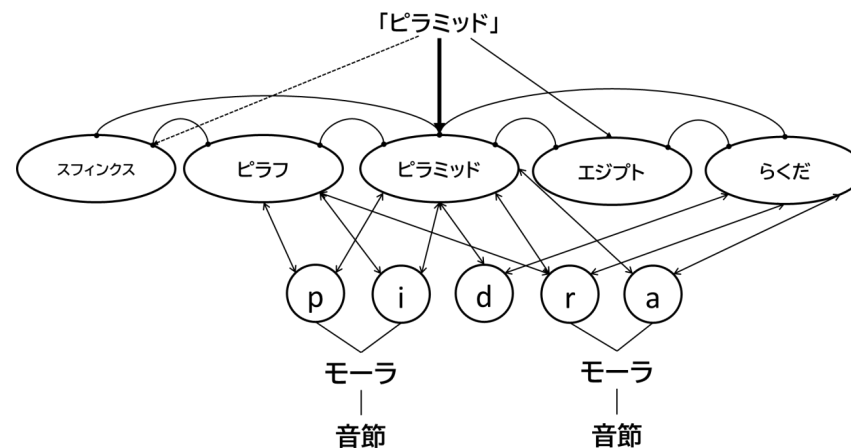
70's



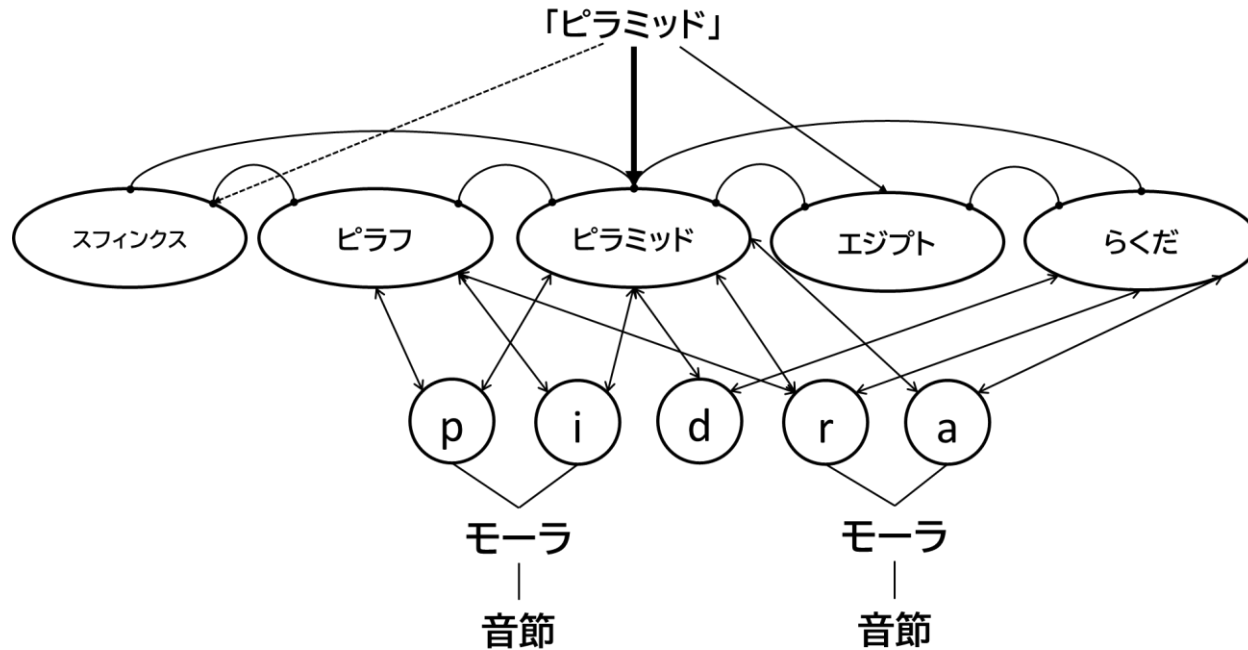
相互活性化モデル 1

1980年代当時注目を集めていたコネクショニズム*の考え方を背景にした
Dell(1986), Stemberger(1985)らのモデル

*たくさんの単純な情報処理要素が、相互に結合して簡単な信号をやりとりするような型のネットワーク状のメカニズムを使って「情報処理」という仕事をさせることで心の構造と機能の解明を目指した動き



相互活性化モデル 2



処理はネットワークの中で双方向的に進む。

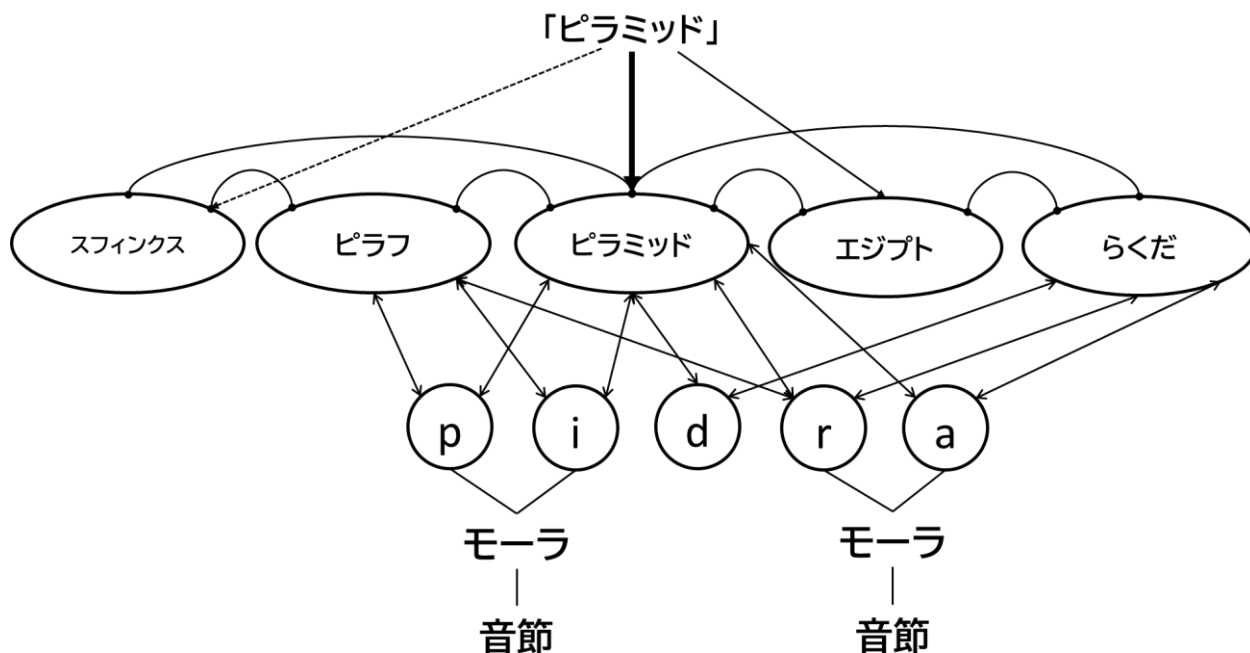
異なるレベル間を活性化が伝播(フィードバックも含む)し、複数のユニット同士で競争が行われる。

他のユニットからの活性化の値の合計がそのユニットの閾値(しきいの値)を越えると次のユニットに活性化が送られる。

活性化が集まる強いユニットがますます強くなるという原理に基づいて最も活性化を受けたユニットが勝者として決定される。

最も活性化を受けたユニットが
勝者として決定される。

相互活性化モデル 3



語彙アクセスの現場が描かれることが多いが、
統語的、音韻的のフレームも別途活性化され、
そのスロットに選ばれた要素が充填されていく。

言い誤りは、

ノイズ(発話のスピード、ユニットの休止レベルのばらつき、反復音素効果)のために、

ライバルのユニットがターゲットのユニットより高い活性化を受けてしまった結果であるとし、

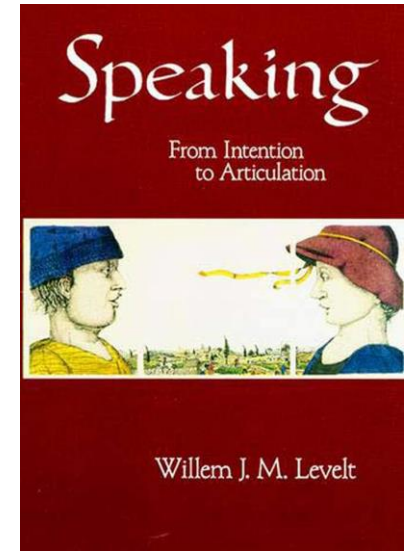
誤りの理由づけがしやすく、発生メカニズムの説明に優れる。

活性化のダイナミズム、ノイズ値の設定を変数として盛り込んでシミュレーションを行うことができる。

原理の単純さゆえ「**トイモデル**」と呼ばれることもあるが、
後の機械学習モデルにつながる特徴を持つ。

Levelt モデル 1

意図から調音までの発話過程の全容解明を目指した
大著「*Speaking*」の冒頭で示された「標準」モデル



90's

Speaking: From Intention to Articulation
Levelt 1993

意図から調音までを捉えた総合的なモデルであったため、
後の研究に向けて、各部門を精緻化していく道筋をつけた。

統語論においてより現実的な文法的符号化を提案した Ferreira(2000)

音韻的符号化から調音までを精緻化したWEAVER++モデル Roelofs(2000)

大規模言語モデル時代の「意図的」とは何か？

Levelt モデル2 全体図

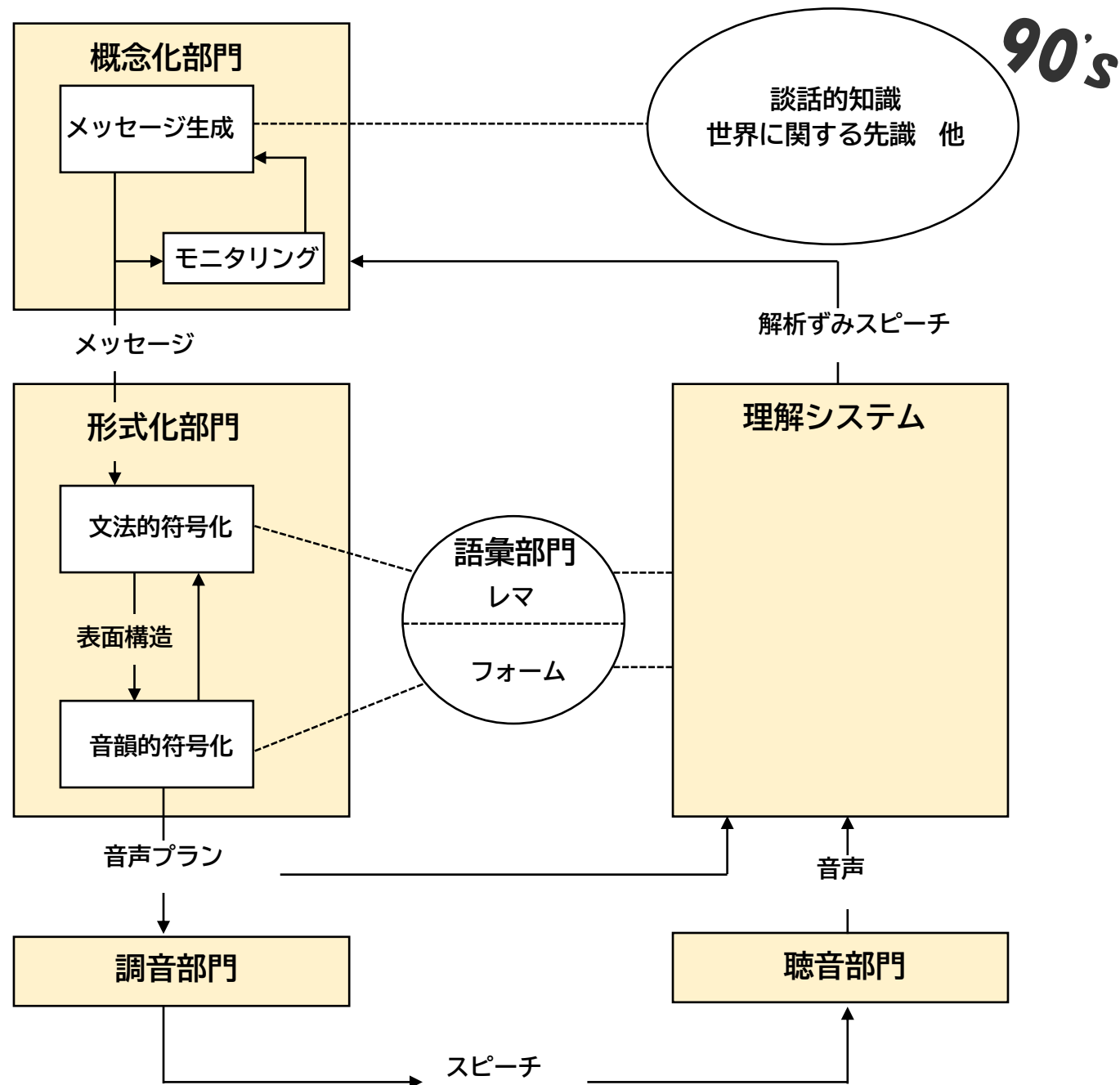
基本的には、
概念化部門、**形式化部門**、**調音部門**
の順に系列的に処理が進む。

語用論的・談話的調整や発話前、発話後の修正のためにフィードバック処理も設けられている。

従来のモデルでは、取り上げられてこなかった**概念レベル**に

- ・語彙概念ツリー
- ・語用論的・談話的調整

を取り入れて
意図の記述を試みている。



Levelt モデル3 全体図

語彙部門

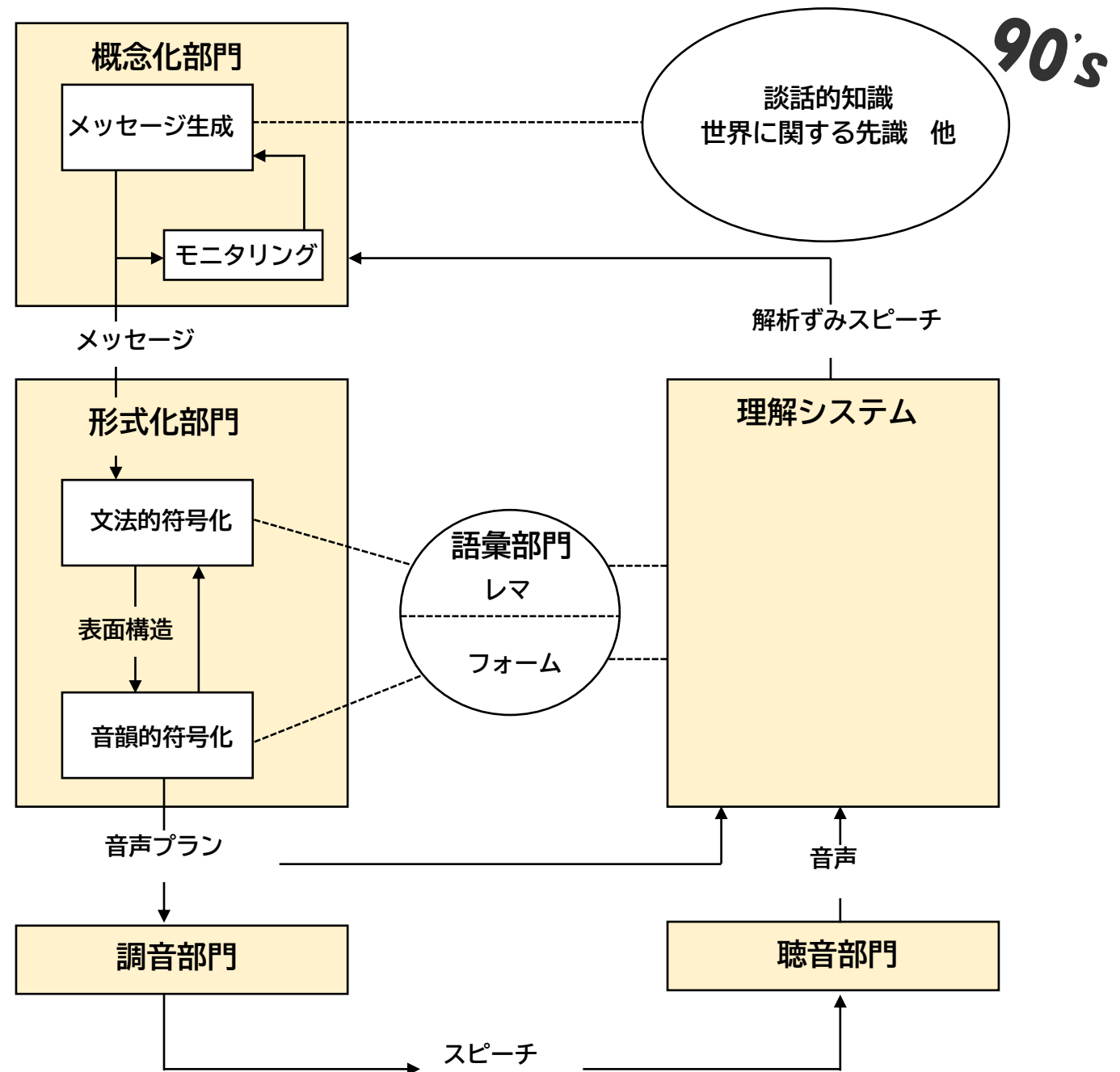
発話・理解の中心に据え、

- ・文法的エントリーとしてのレマ
 - ・形態・音韻的エントリーとしてのフォーム
- それぞれ文法的符号化、音韻的符号化に結びつけて役割を明確にした。

文法的符号化部門

範疇的、統語的手続きに沿って
インクリメンタルに文の組み立てが
進行する「動く」モデル

*シミュレーション等の計算は行えない



認知モデル まとめ

	根拠	デザイン	フィード バック	シミュレー ション	学習	文の組み 立て
Garrett モデル	言い誤り	箱と矢印	×	×	×	△
相互活性化 モデル	言い誤り	ネット ワーク	○	○	×	×
Levelt モデル	実験	箱と矢印	△	×	×	○