2005－SLP－57（13）
2005／7／15

概念学習を効率化するための対話戦略とその獲得

田口 亮　　桂田 浩一　　新田 恒雄
豊橋技術科学大学　工学研究科

対話を通した概念学習では，エージェントが持つ対話調整能力が学習の効率に大きな影響を与える．しかし，有効な対話戦略を如何に与えるかについてはあまり議論されてこなかった．そこで本報告では，2体のエージェントを用いた対話実験を通して，概念学習を効率的に進める上で有効な対話戦略を自動獲得させることを試みる．戦略獲得ではQ学習および Dyna-Q を比較する．両手法の比較実験結果から，Dyna-Q が対話戦略の獲得に有効であることを示す．また，獲得された戦略の評価実験結果から，対話相手の理解状況を効率的に推定する高度な戦略を有効に利用していることが示された．

Automatic Acquisition of Dialog Strategies for Efficient Concept Learning

Ryo Taguchi,　Kouich Katsurada　and　Tsuneo Nitta
Graduate School of Engineering, Toyohashi University of Technology

In the concept learning through human-agent interaction or agent-agent interaction, the efficiency at the learning depends largely on the dialog strategies the agents have. But the procedure to give agents efficient strategies has not been discussed in previous works. This paper describes automatic acquisition of dialog strategies for efficient concept learning applied to the interaction between two agents. In our experiments, Q-Learning algorithm and Dyna-Q algorithm are compared, and the experimental results showed that (1) Dyna-Q agents can acquire more efficient strategies than Q-Learning agents in the first stage of learning, and (2) efficient teaching strategies include sophisticated interaction to confirm counterpart's comprehension level.

１．はじめに

近年，携帯端末，ナビゲーションシステム，ロボットとより自然な対話をすることへの社会的要請が芽生えつつある．人間とエージェントとの対話を考えると，音声認識や表現といったユーザインタフェースとしての側面以上に，エージェントが持つ知識とその使い方に大きな課題が残されている．従来の対話システムでは，対話に利用する知識の全てを開発者が予め想定し，辞書や対話シナリオという形でエージェントに与えてきた．しかし，この手法は(1)開発者の負担が大きい，(2)エージェントが実際に活動する環境や対話相手に接地した知識を与えることが困難である，といった問題がある．こうした背景から近年，人間－エージェント対話を通して，概念をエージェントに自動獲得させる研究が行われ始めた[中川 95, 赤穂 97, 金 00, 新田 02, 小玉 04]．これらの研究では，エージェントが得たセンサ情報と，人間の教示音声の対応関係から概念を学習・獲得することを目指している．また，獲得した概念は，実際の環境および直接の対話相手に接地することを期待している．しかし，一人の人間が教示できる概念数には限界があり，その負担も大きい．そこで我々は，人間との対話を通して幾つかの概念を獲得したエージェン同士が，ネット上のバーチャル空間で対話を行い，お互いが獲得した概念を共有化するシステムを考えている．こうすることによって，一人の人間が教示すべき概念数を削減できるだけでなく，より多くのコミュニティで利用可能な一般的な

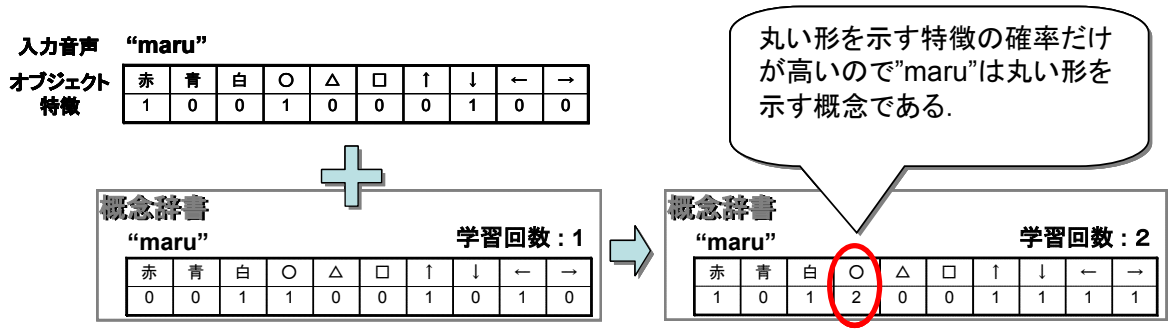

図１：概念学習アルゴリズム

概念も学習することができる.

ところで，対話を通した概念学習を考えると，人間やエージェントが用いる対話戦略が概念学習の効率に大きな影響を与えることは容易に想像できる. 例えば，教示者がランダムに概念を教えるよりも，学習者の理解状況に合わせて教示する方が効率的である. また，学習者が自身の理解状況を教示者に如何に伝えるかによっても効率が変化する. さらに，エージェント同士の対話学習を想定すると，各エージェントが互いの知識量に合わせて適切に教示・質問の役割を切り替えるといった対話調整機能も必要である. しかし，従来の概念学習研究では学習アルゴリズムや獲得／共有化される概念に焦点を置いていたため，有効な対話戦略を如何にエージェントに与えるかという議論は行われてこなかった. 一般に対話戦略は，人間が設計して与えているが，これらは対話相手と対話環境によって大きく異なる. 従って，対話戦略もまた概念と同様，エージェントが自ら獲得し運用するのが望ましい. こうした観点から本報告では，幾つかの概念を獲得したエージェント同士が，対話を通して概念を共有化する過程で，対話学習を効率的に進める戦略を自動獲得させる方法を検討し，比較実験を行う.

２．対話による概念学習

概念学習の先行研究では，音声の一発話に対して複数の属性候補が与えられる1対多学習を対象としている[中川 95, 赤穂 97, 金 00, 小玉 04]. この場合，ウサギの画像に対して「うさぎ」だけでなく，「白い」や「大きい」といった他属性の教示も許可し，画像オブジェクトが持つ特徴と音声との対応関係を学習させることができる. しかし，どの属性に対する教示かは与えられず，エージェントは対象とする属性を自ら判断しなければならない. [赤穂 97, 金 00, 小玉 04]では，属性の判断だけではなく，適切な特徴の範囲（例えば「赤い」という概念が示す色特徴の範囲）も同時に学習させているが，[中川 95]では，各特徴は予めカテゴリ化されていると仮定して実験を行っているため，特徴の範囲は学習の対象外となっている. 本実験では戦略獲得に的を絞るため，最も単純な[中川 95]の方法を概念学習のアルゴリズムとして採用する. 以下，[中川 95]を 2 体のエージェントによる対話学習に拡張した場合の手順を説明する.

対話実験はコンピュータの仮想空間で行われる. この仮想空間には 9 個のオブジェクトがあり，それぞれのオブジェクトは色や形，位置といった複数の視覚特徴を持つ. オブジェクトの視覚特徴は 10 種類（丸，三角，四角，赤，青，白，上，下，左，右）にカテゴリ化され，それぞれの有無を表す 0/1 のベクトルとしてエージェントに渡される. このカテゴリ化されたそれぞれの特徴を以下ではオブジェクト特徴と呼ぶ. 本報告では，このオブジェクト特徴とその特徴を指す音声特徴との対応関係を概念と呼ぶ. 対話はどちらか一方のエージェントが指差しによって話題となるオブジェクトを選択する所から始まる. その後，もう一方のエージェントは指差しで話題を変更し直すか，話題のオブジェクトに関連した獲得済みの概念を 1～4 語で発話（教示発話）する. 教示発話を受け取ったエージェントは，それが未知語の場合には概念辞書（教示された単語と，単語と共起した特徴の頻度を保持）に新規登録する. すでに概念辞書に登録してある単語が発話に含まれている場合には，発話からその部分を切り出し，オブジェクト特徴の頻度を更新する（図1参照）. もし未知語と既知語が同時に教示された場合には，既知語部分を切り出した後，残りの部分を未知語として登録する. そして，出現確率（頻度÷

学習回数)が0．9以上になる特徴が一つだけある場合に，その特徴と音声の対応関係を概念として獲得する．

このタスクには以下の二つの問題がある．

① 発話された単語がそれぞれどのオブジェクト特徴のことを指しているかについての情報は相手に与えられない．

② 発話は単語毎に区切られておらず，連続した音声として与えられるため，連続して 2 語以上の未知語を教示すると，1 つの未知語として受け取られてしまう．

①の問題から，例えば「あか」という概念を獲得させるために教示者は，赤い丸や赤い四角などの複数の赤いオブジェクトに対して「あか」と教示し，それが色の概念であることを確率的に学習させる必要がある．複数単語による教示はそれを効率的に行うために有効である．しかし，②の問題があるため単純に「できるだけ多く発話する」だけでは効率的な対話は実現しない．本実験ではこうした教示のための戦略に加え，「いつ聞き返すべきか」や「どのオブジェクトを話題にするか」といった学習者側の戦略も同時に獲得させる．尚，本実験では対話戦略を獲得する手法の確立に的を絞るため，音声特徴には音素単位のシンボル列を用いた．また，未知語を正確に聞き取ることは，既知の単語を聞き取るよりも困難であると仮定し，未知語を受け取った IA は 1 音素当たり 0.1 の確率で認識誤り(ランダムに音素を変化)を起こすとした．

３．対話戦略の獲得方法

本報告では，一般的な強化学習の手法であるQ学習[Sutton 98]と，強化学習および状態空間プランニングを併用した Dyna-Q アルゴリズム[Sutton 98]の二つ対話戦略の獲得実験から比較する．

3.1. Q学習

強化学習とは環境を予めモデル化することなく，エージェントが行動した際に得られる報酬を元に取るべき行動を学習していくアルゴリズムである．エージェントが環境を認識した結果を状態と呼ぶ．一般に学習後は，各状態における最適な行動が学習されるため，条件反射的な素早い意思決定を可能にする．Q学習は強化学習のアルゴリズムとして一般的に広く利用されている．

Q 学習の目的は状態－行動ペアの価値(Q 値と呼ばれる)を推定することである．状態 s で行動 a を取る際のQ値はQ(s,a)と表される．エージェントが行動するたびにQ値は更新されていき，Q値の収束後は各状態において最も高いQ値を持つ行動を選ぶことが最適な戦略になる．

状態sで行動aをとり，報酬rを得て，状態s’に遷移した場合のQ値の更新式を以下に示す．但し，αは学習率，γは割引率を表す．

$$Q(s,a)=(1-\alpha)Q(s,a)+\alpha(r+\gamma \max Q(s',a))$$

3.2. Dyna-Q

強化学習は環境を予めモデル化することなく，エージェントと環境との相互作用を通して条件反射的な行動を学習していく．一方，プランニングとは，与えられた環境のモデルから戦略を導出，または改善するための手法であり，熟考型の意思決定を実現する．プランニングは大別すると，記号論理をベースとしたプラン空間プランニングと，動的計画法などに代表される状態空間プランニングにわけられる．後者は強化学習との親和性が高く，強化学習とプランニングを統一的に扱うアーキテクチャが提案されている．本報告ではその一つであるDyna-Qを利用して実験を行う．

Dyna-Qのアーキテクチャを図2に示す．Dyna-Qエージェントは，環境とのインタラクションで得た経験(図中：実際の経験)から強化学習と環境モデルの学習を行う．またオフラインでは，学習した環境モデルから得られるシミュレーション上の経験を利用しプランニングを行う．強化学習とプランニングには先述の Q 学習が用いられる．すなわち，Q学習に入力する経験を切り替えることで，強化学習とプランニングを併用した戦略の獲得ができる．そのため，Q学習のようなオンライン学習のみの手法よりも，少ない経験で戦略獲得が実現できる．但し，オフラインの学習が適切に実行されるためには，正確な環境モデルを学習しておく必要がある．最も単純な環境モデルの学習方法は，各状態行動対における次状態とその頻度，得られた報酬を全て保持し，状態遷移確率と平均獲得報酬を計算することである．しかし，本実験のようなマルチエージェント学習問題では，相手が戦略を学習することによって環境が変化してしまう．そこで本実験では，状態遷移確率および報酬を時間と共に忘却するようにした．

次節では，Q学習および Dyna-Q で用いるエージェントの状態，行動，報酬について説明する．

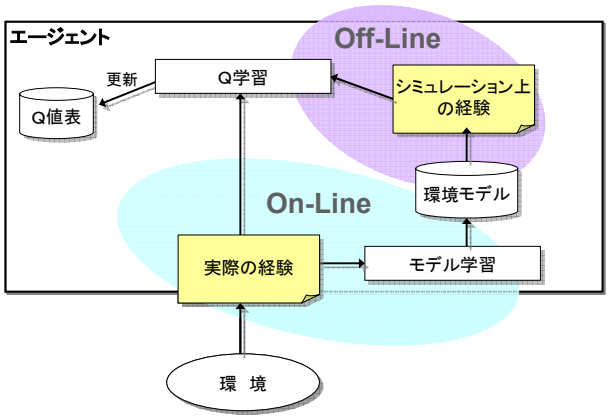

図２：Dyna-Q アーキテクチャ

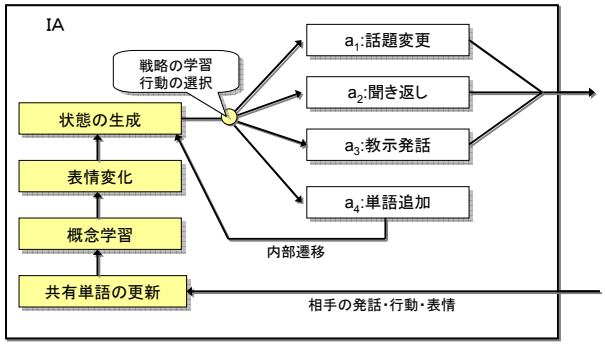

図３：エージェントの行動

3.3. 状態・行動・報酬

(1) 行動

IA は以下の4つの行動を持つ.

・話題変更:ランダムにオブジェクトを選択し指差す.
・聞き返し :相手の教示発話を繰り返す.
・単語追加:話題となるオブジェクトに関連する単語を一つ発話レジスタに追加する.
・教示発話:発話レジスタの内容を発話(1～4 語)する.

対話はどちらかのエージェントが「話題変更」することによって始まる. その後は, 学習中の戦略に従って上記の 4 つの行動のどれかを選択し実行していく.「話題変更」,「聞き返し」,「教示発話」を実行した場合は相手に行動の権利が移るが,「単語追加」は話題に関連する単語を全て追加するまで(最大 4 語)繰り返し実行することができる. また, 相手のエージェントが「聞き返し」や「教示発話」をした場合, その中で正確に発話された単語を「両者で共有された単語」(共有単語)として共有単語メモリに保持する. エージェントの行動の流れを図 3 に示す.

(2) 表情変化

各エージェントは快, 不快, 平常の感情(生得的な内部状態)を持ち, 以下の規則に従って変化する. これらの感情は表情モダリティを介して他のエージェントに伝えられ, 強化学習および環境モデルの状態と報酬に利用される.

・快 : 新たな概念を獲得した場合
　　共有単語が増加した場合
・ 平常 : 快でも不快でもない場合
・ 不快 : 相手の発話に未知語が含まれている場合

(3) 状態

各エージェントは, 相手の行動, それによって変化した自身の表情, 今何を発話しようとしているのか, といった情報を用いて状態を生成する. 具体的には, 以下の7次元で状態を表現する.

・相手の表情(快, 平常, 不快)
・相手の行動(話題変更, 聞き返し, 教示発話)
・自分の表情(快, 平常, 不快)
・自分の獲得概念数(0～10)
・話題となるオブジェクトに関する既知の概念数(0～4)
・発話レジスタ内の単語数(0～4)
・発話レジスタ内の未共有単語数(0～4)

(4) 報酬

協調的な対話戦略を獲得させるために, 報酬は両エージェントとも共通とした. 具体的には, エージェントの感情が快になった場合に + 10, 平常／不快の場合に - 1, 単語追加時の内部遷移の場合に 0 とし, その報酬と相手の表情から算出される報酬とを足したものを報酬として利用した.

4. 対話戦略の獲得実験

前節で設計した 2 体のエージェントによる対話学習を通した戦略の獲得実験を行う.

4.1. 実験条件

実験に使用した概念はオブジェクト特徴に対応した 10 個とする. 両エージェントの概念が 10 個以上になるか, 対話が 100 ターンを超えるまでを 1 試行とし, 各試行の最初に各エージェントの持つ概念を初期化する. 但しこの時, Q 値表(戦略の学習結果)は初期化せずに保持される. 試行を1万回繰り返し, 両エージェント同時に対話戦略を獲得させる.

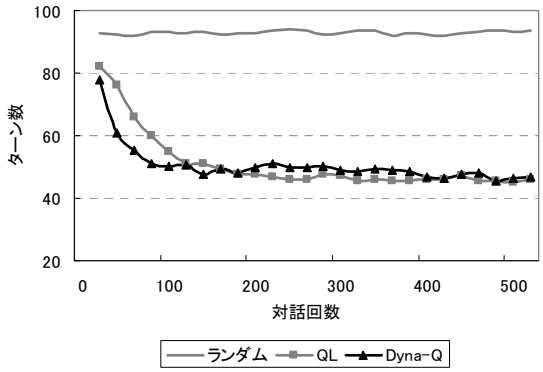

図４：Q学習と Dyna-Q の比較（対話回数 500）

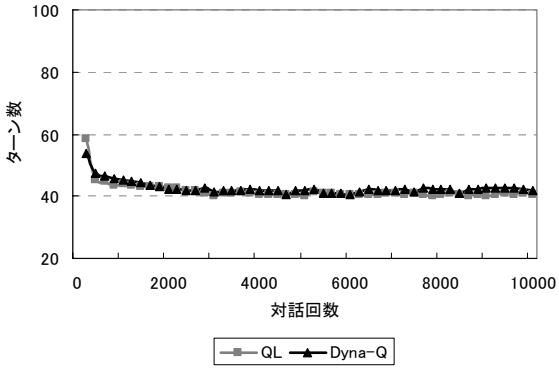

図５：Q学習と Dyna-Q の比較（対話回数 10,000）

本報告では 2 種類の条件で対話実験を行う．条件 1 では試行毎にエージェント A の初期概念を 10，エージェント B の初期概念を 0 に初期化するという条件の下，対話戦略を獲得させる．条件 2 では実験条件をより現実に近づけるため，エージェント B に与える初期概念を未知とした場合，すなわち 0～9 個の概念をランダムに与えた場合の実験を行う．Q 学習の学習率 α は 0.1～0.001 へと減少させる．割引率 γ は 0.9 と設定した．学習時の行動選択には ε グリーディ方策（ε = 0.1）を用いている．Dyna-Q によるオフラインでのプランニングは 10 試行毎に 1 万ステップ（行動）行う．

4.2. 実験結果

(1) Q学習と Dyna-Q の比較

実験条件 1 におけるQ学習と Dyna-Q の学習曲線（20 回の平均）を図 4，5 に示す．図の横軸は試行回数，縦軸はエージェント B が 9 個の概念を獲得するまでにかかった時間（ターン数）である．比較のためにランダムで行動した場合の結果も載せる．なお，プランニングに要した時間（ステップ数）は載せていない．図 4 から学習初期において Dyna-Q が，Q 学習よりも効率的な対話戦略を獲得できることが解る．また，図 5 の結果をみると，学習の後半では Dyna-Q とQ学習で獲得される戦略の性能がほぼ同等となっている．この結果は，Dyna-Q を利用することでより早く戦略が獲得できることを示している．

(2) 初期概念数の違いが獲得戦略に与える影響

1万回の試行によって得られた獲得戦略(最も Q 値の高かったもの)を観察すると，以下のようなものであった．全ての概念を保持しているエージェント A は教示発話を行う教示戦略を獲得した．一方，エージェント B は主に話題変更と聞き返しを行う質問戦略を獲得した．この結果は，与える初期概念数の違いによってエージェントの役割が適切に分化したことを示す．尚，先述の条件 1 で獲得された戦略をそれぞれ教示戦略 1，質問戦略 1，条件 2 で獲得された戦略を教示戦略 2，質問戦略 2 と呼ぶ．

質問戦略 1 と 2 は，ともに「教示発話に未知語がなく，新たな概念獲得もなかった場合，話題を変更する」という戦略を持っていた．この戦略は概念学習が進まなくなった場合に話題を変更することができ，概念学習を効率化する．また，質問戦略 2 では獲得概念数が増えると教示発話を行うという戦略が見られた．他方，教示戦略は 1 と 2 で異なっており，教示戦略 1 は，「発話に含まれる未共有単語（双方が共有していない語）が 1 語以下になるように教示する」という内容になっていたのに対して，教示戦略 2 では「相手が話題変更した場合には一度にできるだけ多くの概念を発話し，それ以外の場合では教示戦略 1 と同じ戦略を取る」という内容になっていた．概念学習において，多くの未知語が含まれる発話は，それぞれの未知語を正確に切り出すことが困難になる．従って，できるだけ多くの概念を発話することは，相手の知識量によって受け取られ方が変わる不確実性の高い教示と言える．図 6 は，対話学習を終えた後，二つの獲得戦略を用いて評価実験を行った結果を示している．

1万回の試行の後，各エージェントが獲得した戦略を用いて，評価実験を行った．ここでは，戦略の学習は行わずに，エージェント B に与える初期概念数を 0，3，6

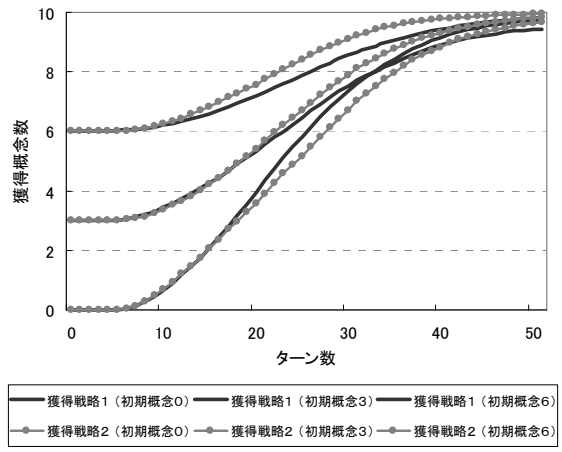

図 6：獲得戦略の比較

とした場合の対話学習を行った． 獲得戦略 2 は，エージェント B の初期概念が 3 個以上の場合に，獲得戦略 1 よりも効率的であった．これは教示戦略 2 が，不確実性の高い教示をすることで相手にとっての未知語を効率よく確認することができ，さらに未知語が見付かった場合には教示戦略 1 と同様の正確な教示ができるためである．相手の知識量が解らないという曖昧性を含んだ環境においては，不確実性の高い教示をうまく利用して，対話相手が獲得している概念を効率的に推定していく戦略が有効であることを示している．

4.3. 考察

Grice は意味が一義的に限定される発話をするための4つの格率（量，質，関係，様態）を挙げている[石崎 01]．本実験で獲得された教示戦略1はその一つである量の格率を厳守する戦略になっており，教示戦略 2 は量の格率を一部違反する戦略になっていた．従来の対話エージェントの研究では，曖昧性を含む発話はタスク達成の妨げになると考えられており，開発者は意識的にまたは無意識に先の格率を厳守するように対話戦略を設計してきた．しかし，人間同士の対話ではそういった格率を厳守するような発話ばかりではなく，あえて格率に違反した発話（意味が一義的に限定されない不確実性の高い発話）を行い，相手の反応を期待するという教示戦略 2 と同様の戦略が日常的に使用されている．そしてそれは本実験の結果が示すように，曖昧性を含む環境で対話を効率的に進めるためには有効な戦略である．しかし，このような不確実な発話を許容する戦略は，格率を厳守するように設計された戦略よりも複雑になるため，従来のようにヒューリスティックで設計することは困難である．本報告で提案する対話戦略の獲得は，そういった高度な対話制御を実現するための方法として有効である．

5．まとめ

本報告では，2体のエージェントを用いた対話実験を通して，概念学習を効率的に進めるために有効な対話戦略を自動獲得させる実験を行った．戦略獲得の方法にはQ学習および Dyna-Q を用いた．対話実験の結果から，(1) Dyna-Q を用いることで少ない経験で有効な対話戦略が獲得できること，(2) 対話相手の理解状況を効率的に推定する高度な戦略を有効に利用していることが示された今後は，エージェントに音声認識・合成を組み込み，人間との対話を対象とした実験を行っていきたい．

参考文献

[赤穂 97] 赤穂，速水，長谷川，吉村，麻生: EM 法を用いた複数情報源からの概念獲得,信学会論文誌，Vol.J80-A pp.1546-1553 ，1997.

[金 00] 金，岩橋: 知覚情報の統合に基づく言語音声単位の獲得アルゴリズム，信学技報，TL200-21，pp.9-16，2000.

[小玉 04] 小玉，田口，桂田，岡部，新田: オンライン学習による Infant Agent のための効率的概念獲得，人工知能学会全国大会，2004，3F3-03.

[中川 95] 中川，升方: 視聴覚情報の統合化に基づく概念と文法の獲得システム，人工知能学会，Vol.10，No.4，pp.619-627，1995.

[新田 02] 新田，越坂，桂田: Infant Agents 間での対話による概念知識獲得，人工知能学会全国大会，2002，1A1-07

[Watkins 92] C.J.C.H. Watkins，P.Dayan: Q-learning, Machine Learning 8 , pp.279-292 , 1992.

[Sutton 98] R.S. Sutton，A.G.Barto: Reinforcement Learning，MIT Press，1998（三上ほか 訳：強化学習，森北出版，2000）.

[石崎 01] 石崎，伝康: 言語と計算3 談話と対話，東京大学出版会，pp.13-31 ，2001.