

講義「情報理論」

第1回 情報理論とは

情報理工学専攻 情報知識ネットワーク研究室
喜田拓也

今日の内容

- 1.1 情報が伝わるとは
- 1.2 情報伝達のモデル
- 1.3 情報伝達の理論
- 1.4 情報理論の歴史
- 1.5 現代の情報理論

情報とは？

そもそも情報ってなに？

広辞苑を引いてみました

[株式会社岩波書店 広辞苑第五版]

- ① あることがらについてのしらせ。「極秘—」
- ② 判断を下したり行動を起したりするために必要な、種々の媒体を介しての知識。「—が不足している」

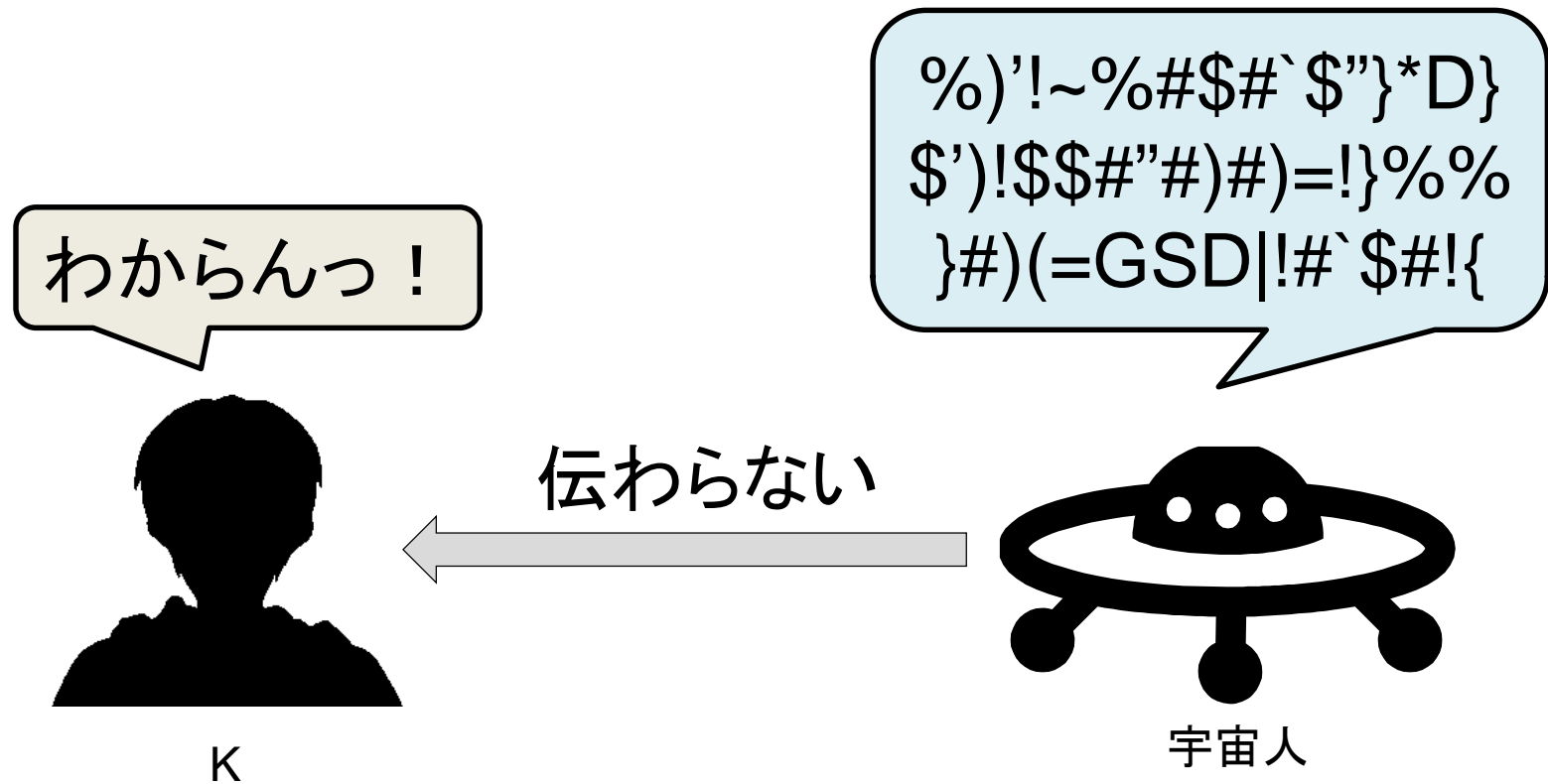
要するに？

伝達されうる知識

片方の(脳内)世界に存在する事象・概念を、
他方の世界へ伝えるもの

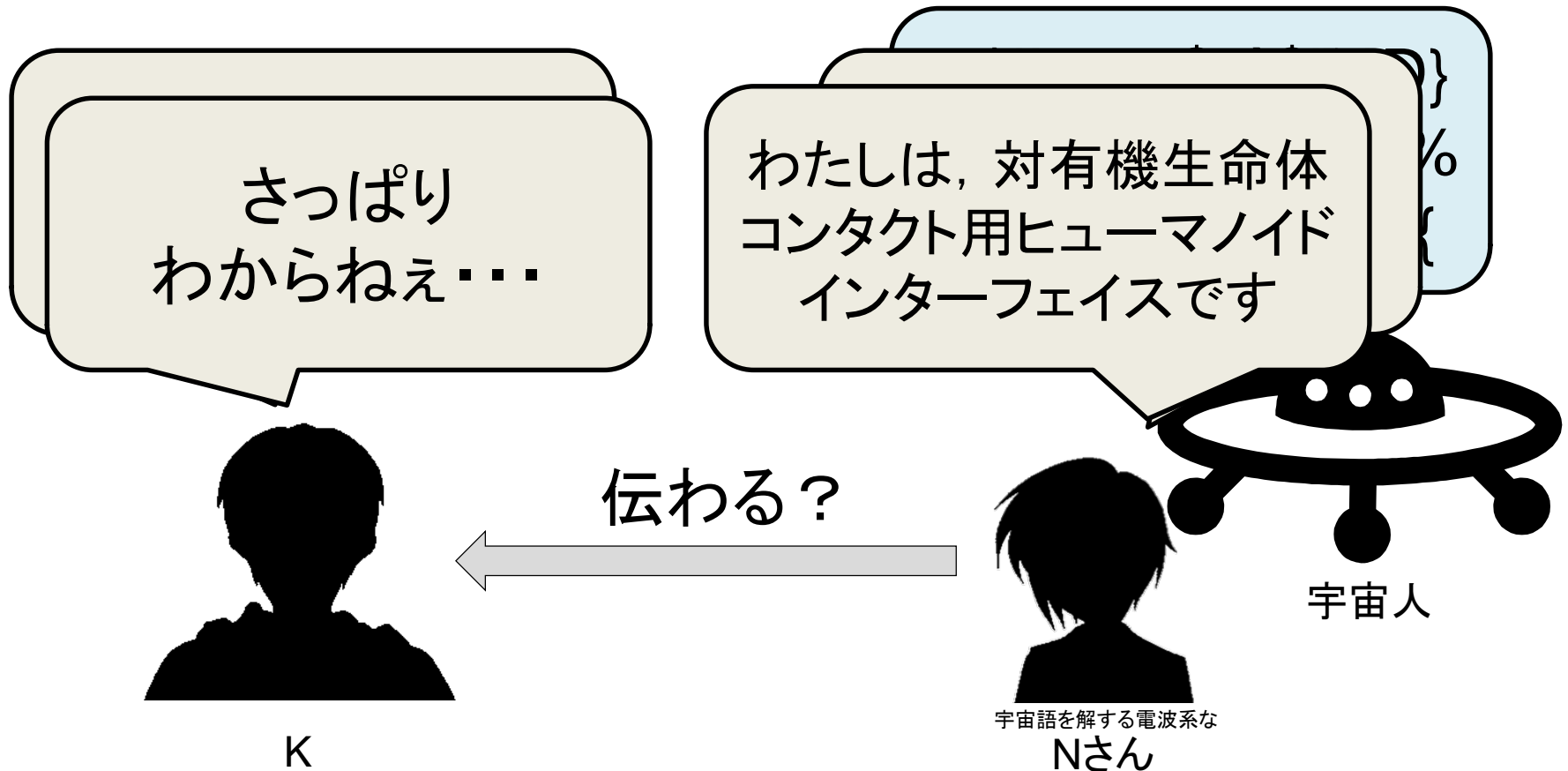
(°Д°)ハア？

情報が伝わるとは



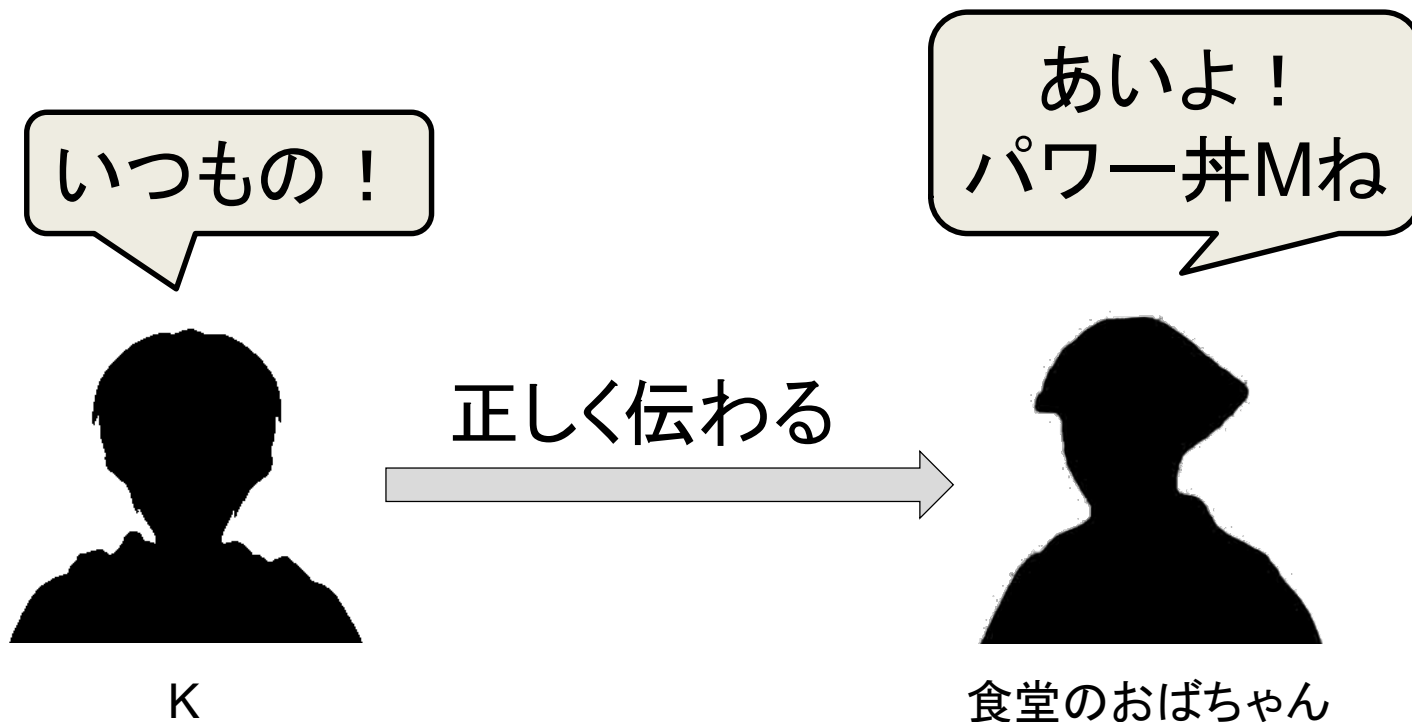
双方で**使う言葉が同じ**でないとは何も伝わらない

情報が伝わるとは



受け手の世界 (持っている知識) が異なると、うまく伝わらない

情報が伝わるとは



双方が知識を共有していれば、**短い言葉(データ)**で伝えられる

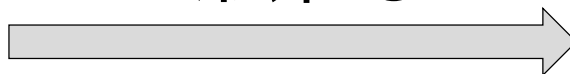
情報が伝わるとは

気持ちは大盛りで
お願いします！



K

正しく伝わる？



あいよ！
気持ちは大盛りね



食堂のおばちゃん

受け手側の世界に**変化**がなければ、情報が伝わったとは言えない
(意味・意図を正確に伝えることは本当に難しい)

情報伝達の本質

情報伝達の本質は、**受け手の知識の変化**である

受け手はデータを受け取ることにより知識が増える
この増えた知識こそが、伝わった**情報**なのである

じゃあ知識が変化するって、
具体的にどういうこと？



K

統計モデルで
表現される



N

とある統計モデルの例

今日の天気について，朝起きてカーテンを開けるまで，あなたが知っている知識はない

ところが，どのような天気が多いかについては，手元に**統計データ**があるとする（今あなたが持っている知識）

表1.1：札幌の1月の天気出現率

晴	曇	雨	雪
5.5%	1.2%	0.2%	93.1%

カーテンを開けたら，今日の天気についてあなたは100%の知識を得たことになる

問1.1をやってみよう！

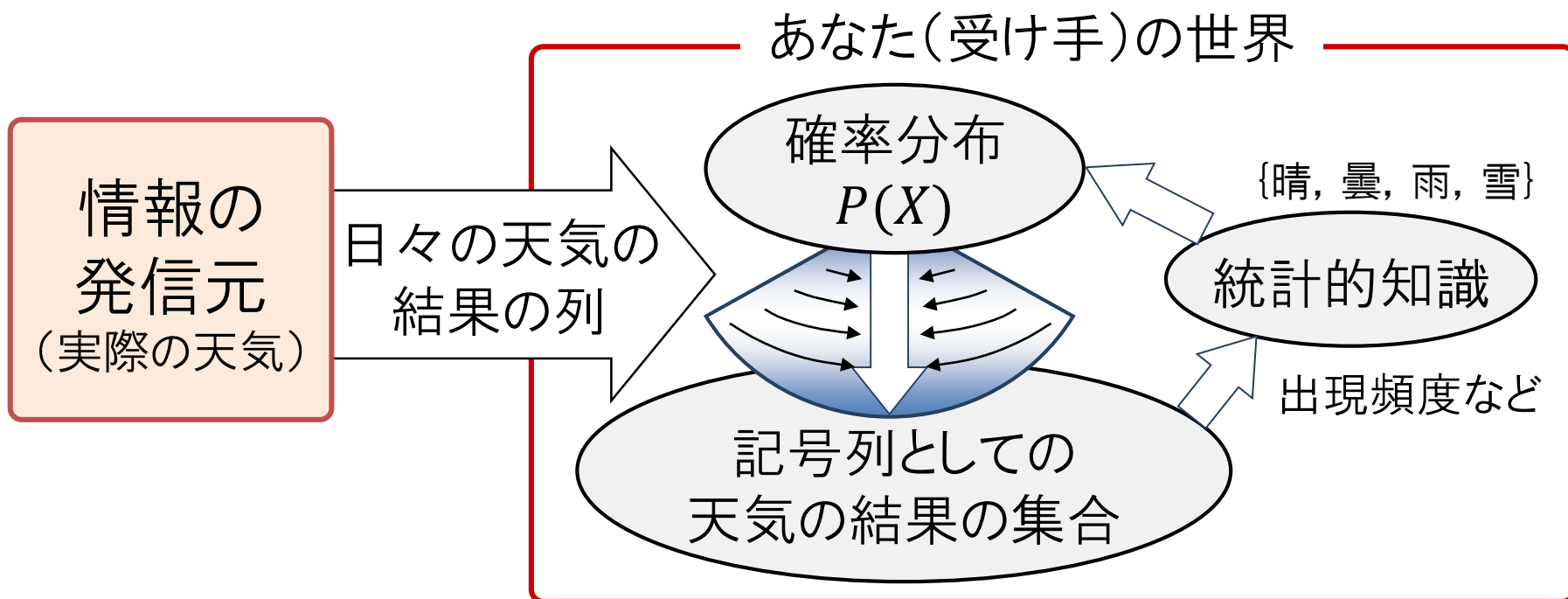
記号列の確率分布変化としての情報伝達

情報を「**記号の列**」だと割り切って取り扱う

その記号列の**意味内容には立ち入らない**

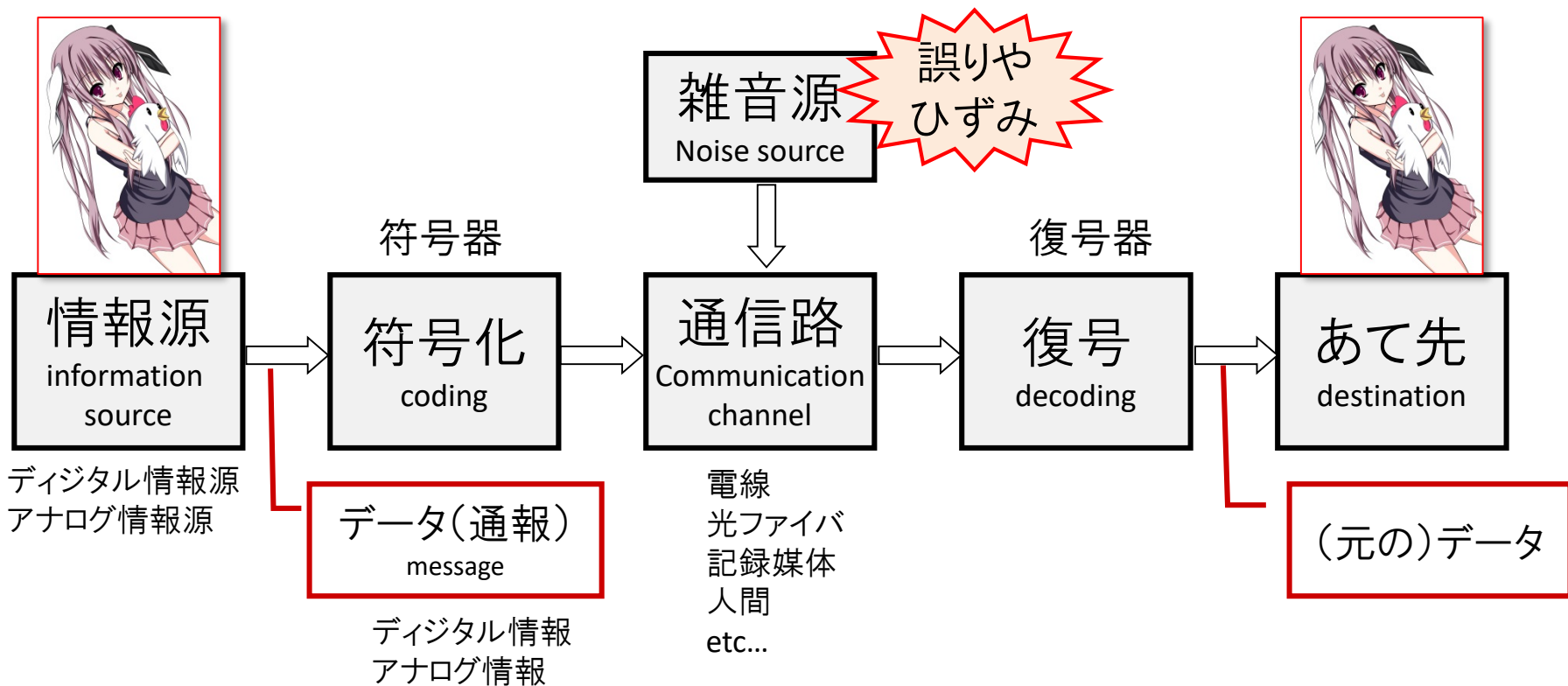
送信される情報は記号 {晴, 曇, 雨, 雪} の列

あなた(受け手)の世界は, 記号(列)の集合とその統計的知識



ちよつと休憩

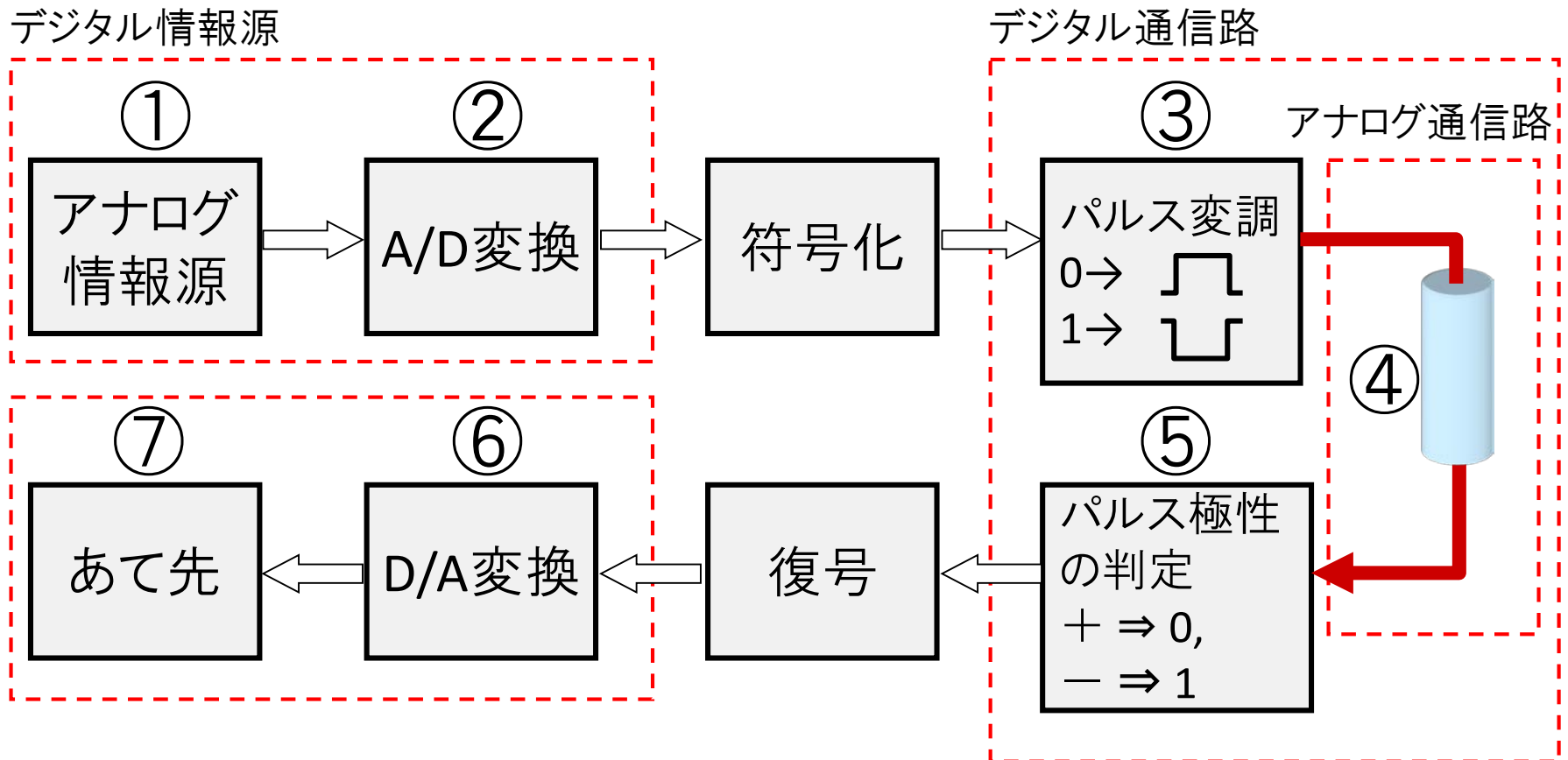
通信システムのモデル



情報源が発する**通報**を、**通信路**を通して宛先へ伝達する際、
情報伝達の**効率**や**信頼性**について考える

実際の情報源・通信路とモデルの対応

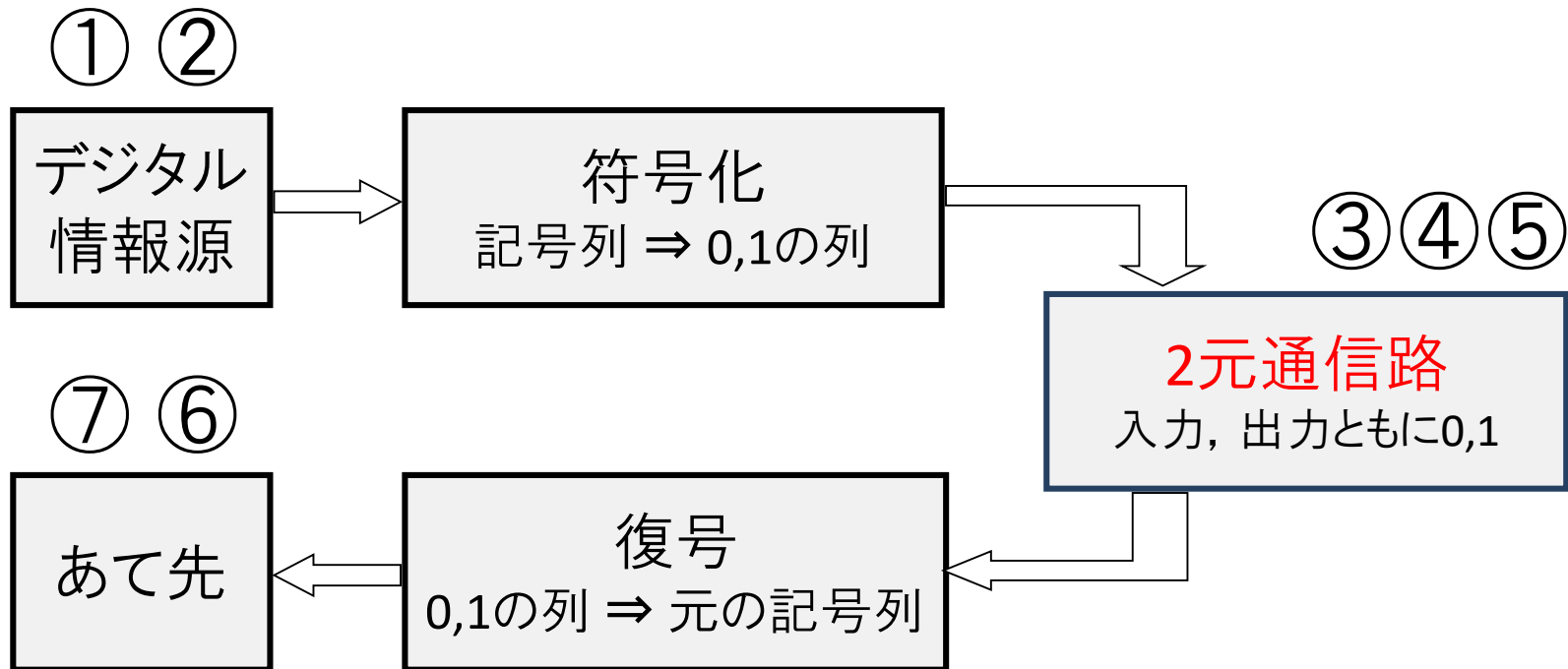
実際の通信システムを考えると、前述のモデルにどのように当てはめるかは、一意には定まらない



本科目の主題：デジタル情報源の符号化

情報源は記号列を出力する(デジタル情報源)

通信路は記号列を伝搬する(デジタル通信路)



情報理論の問題

問1.2) 天気の情報に2元通信路で余所へ送りたい。通信路を使用する際には、送られる記号数に応じて課金されるので、できるだけ送る記号数を減らしたい。どうすればいい？

問1.3) 2元通信路では、困ったことに、1記号毎に 10^{-3} の確率で0と1が反転する。どうにかして、誤りを少なく記号を送りたい。どうすればいい？

表1.2

情報源記号	確率	C1	C2
晴	0.055	0 0	1 0
曇	0.012	0 1	1 1 0
雨	0.002	1 0	1 1 1 0
雪	0.931	1 1	0

符号

符号語

符号アルファベット = {0,1} :
2元符号 (q個あったらq元符号)

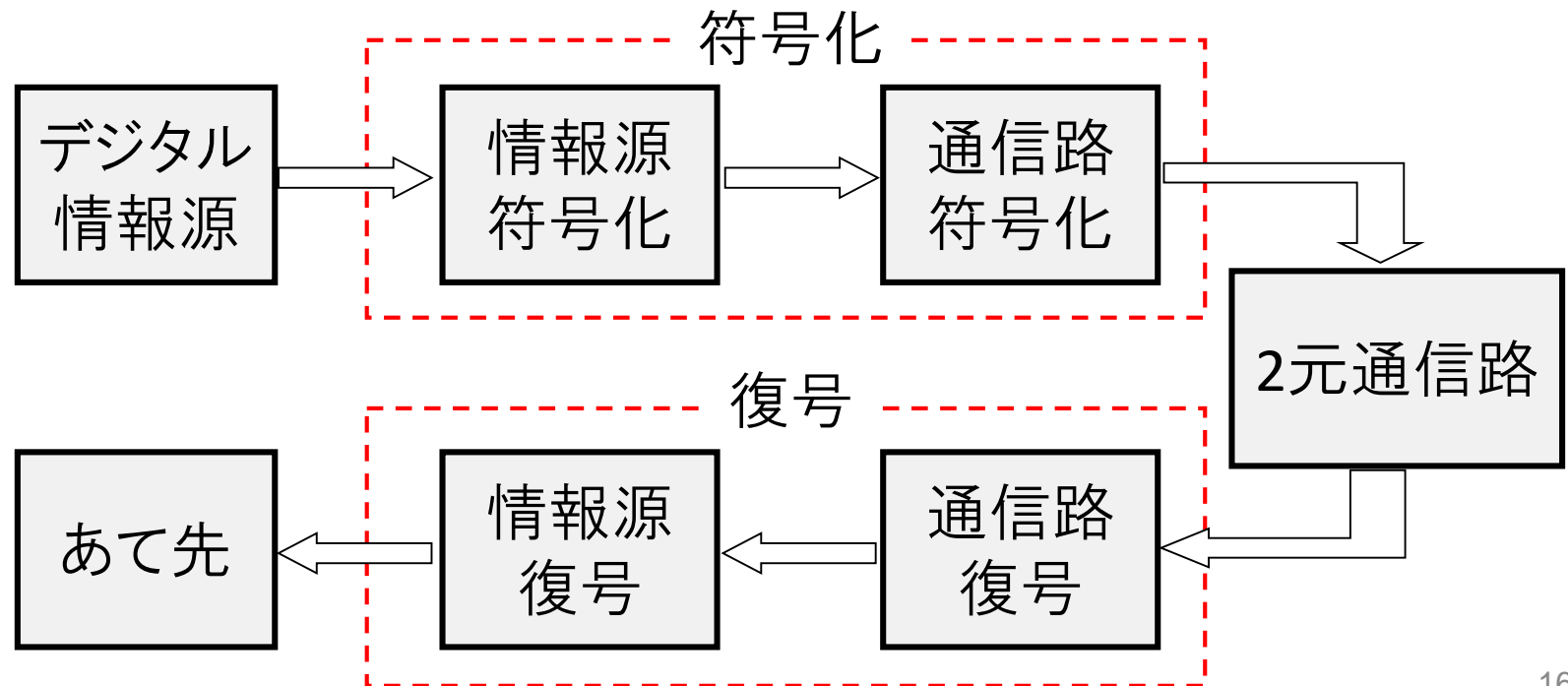
情報理論の問題

次の二つを達成する**具体的な符号化の方法**とそれによってどこまで改善できるかの**理論的限界**を探る

通信路使用の**効率**(efficiency)の向上

信頼性(reliability)の向上

符号化部分を二つに分けて考える！



情報源符号化の問題

【問題1】

できるだけよい情報源符号化と復号の方法を見出す

1 情報源記号あたりの符号系列の長さの平均値(平均符号長という)ができるだけ小さいことが望ましい

装置化が簡単で, 符号化・復号による遅延が小さいほどよい

【問題2】

情報源符号化の限界を知ること

1 情報源記号あたりの平均符号長をどこまで小さくできるか？

※ 可逆符号化(情報無損失符号化)と非可逆符号化(情報損失符号化)がある

通信路符号化の問題

【問題3】

できるだけよい通信路符号化と復号の方法を見出す

付け加えた冗長性を信頼性向上に可能な限り有効に活用できる符号化が望ましい

復号した後の記号の誤り率・冗長度の最小化

【問題4】

通信路符号化の限界を知ること

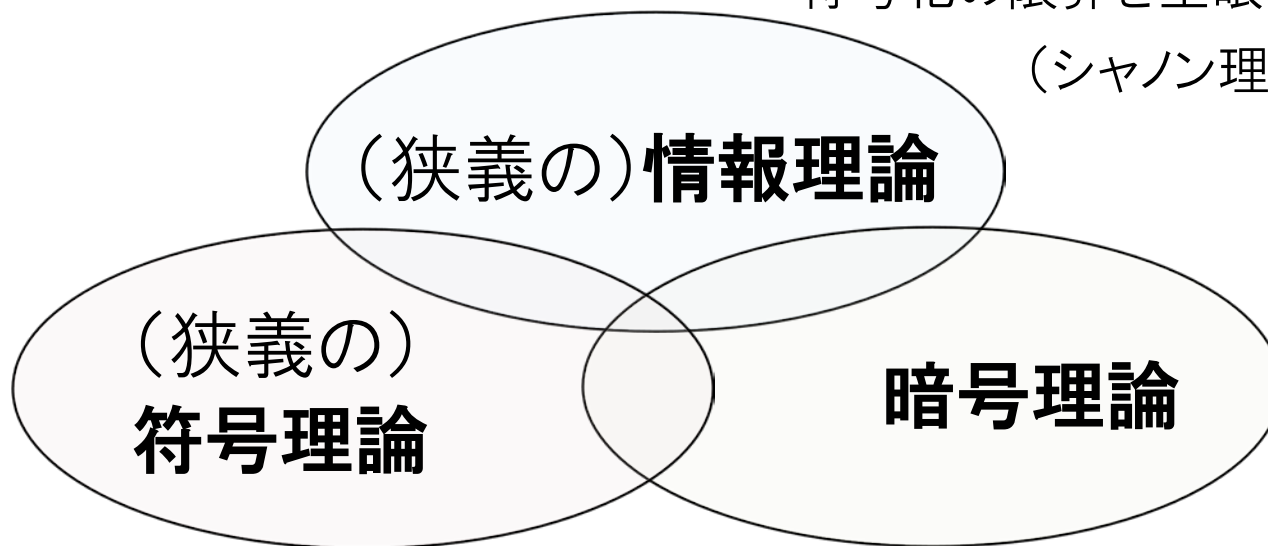
復号後の記号の誤り率をある値以下に抑えたとき、付加すべき冗長度をどこまで小さくできるか？

情報理論の位置づけ

Claude Elwood Shannon(クロード・E・シャノン)博士の論文

C. E. Shannon, "A mathematical theory of communication,"
Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, July and October, 1948.

符号化の限界を主眼とする理論
(シャノン理論ともいう)



通信路符号化の具体的構成法および
符号化・復号法を主眼とする理論

安全な情報伝達の理論

今日のまとめ

情報伝達のモデル化について学んだ

送信側と受信側で、共通の言語・世界のモデルが必要

記号列の統計的性質による情報伝達モデル

情報源→情報源符号化→通信路符号化→通信路→
→通信路復号化→情報源復号化→あて先

情報理論が取り組む4つの問題

【問題1】できるだけよい情報源符号化法(復号化法)を見出すこと

【問題2】情報源符号化の限界を知ること

【問題3】できるだけよい通信路符号化法(復号化法)を見出すこと

【問題4】通信路符号化の限界を知ること

情報理論の分野について

次回テーマ:「情報量とエントロピー」