

# Théorie de l'information

## Cours 0 - Introduction

Laurent Oudre  
laurent.oudre@univ-paris13.fr

Université Paris 13, Institut Galilée  
Master Ingénierie et Innovations en Images et Réseaux - 1<sup>ère</sup> année  
2017-2018

## Sommaire

Organisation du cours

Qu'est-ce que la théorie de l'information ?

## Sommaire

Organisation du cours

Qu'est-ce que la théorie de l'information ?

Un père fondateur : Claude Shannon  
Concept d'information  
Tentative de définition

## Quelques informations pratiques

- ▶ 9 séances de CM/TD
- ▶ Présence obligatoire à toutes les séances, absences comptabilisées
- ▶ 1 séance de révisions
- ▶ Examen : 3h, documents non autorisés (sauf formulaire)

<http://www.laurentoudre.fr/ti.html>

## Sommaire

### Organisation du cours

Qu'est-ce que la théorie de l'information ?  
 Un père fondateur : Claude Shannon  
 Concept d'information  
 Tentative de définition

## Un peu d'histoire

Claude Shannon (1916 - 2001) est un ingénieur électricien et un mathématicien américain.



- ▶ Shannon comprend que toute donnée, même la voix ou les images, peut se transmettre à l'aide d'une suite de 0 et de 1 (les bits), ouvrant la voie aux communications numériques et non plus analogiques.
- ▶ Il étudie les propriétés théoriques des messages numériques binaires et de leur transmission, ce qui donne lieu à des applications en cryptographie, en compression de données mais également pour la correction d'erreurs de transmission.
- ▶ Il s'intéresse également à l'intelligence artificielle et conçoit le premier ordinateur qui joue aux échecs.
- ▶ Son article fondateur, publié en 1948, fut une avancée conceptuelle parmi les plus importantes du XXe siècle, et a permis le développement, l'étude et le perfectionnement des systèmes de communication que nous utilisons chaque jour (Internet, téléphonie, satellite...)

## Théorie de l'information

Il n'y a pas une, mais plusieurs théories de l'information ! Le concept d'information a été introduit et utilisé dans différents domaines :

- ▶ Télécommunications : Shannon
- ▶ Physique et thermodynamique : Boltzmann
- ▶ Informatique et algorithmique : Kolmogorov, Solomonov et Chaitin
- ▶ Economie : Ozga et Stigler
- ▶ ...

La plus utilisée et celle que nous allons étudier dans ce cours est la **théorie de l'information de Shannon**.

## Qu'est-ce que l'information ?

- ▶ Notion a priori assez floue !
- ▶ Exemple habituel : le journal télévisé. Dans ce contexte, l'information peut être vue comme un message transmis d'une source (le présentateur) à un destinataire (le téléspectateur) par l'intermédiaire d'un média (la télévision)
- ▶ L'un des buts de Shannon est de quantifier avec une valeur numérique la quantité d'information contenue dans un message (par exemple ici, dans une annonce faite par le présentateur télé)

## Qu'est-ce que l'information ?

Considérons deux annonces :

A - La Terre va exploser demain  
B - Il va pleuvoir en Bretagne demain

- ▶ Laquelle contient le plus d'information ? Clairement l'annonce A !
- ▶ Pourquoi ? Parce que c'est une information qui nous surprend, et qui nous semble impossible ou du moins très peu probable
- ▶ On voudrait donc que A soit associée à une quantité d'information très élevée et B à une quantité d'information faible

## Tentative de définition

Selon Wikipédia,

*La théorie de l'information, sans précision, est le nom usuel désignant la théorie de l'information de Shannon, est une théorie **probabiliste** permettant de **quantifier** le contenu moyen en **information** d'un ensemble de messages, dont le **codage** informatique satisfait une **distribution statistique** précise. Ce domaine trouve son origine scientifique avec Claude Shannon qui en est le père fondateur avec son article A MATHEMATICAL THEORY OF COMMUNICATIONS publié en 1948.*

- ▶ Notion d'information et de quantification de l'information
- ▶ Vision probabiliste de l'information
- ▶ Lien avec le codage

## Qu'est-ce que l'information ?

Considérons maintenant les deux annonces :

A - La Terre va exploser demain  
C - Demain nous allons tous devenir bleus

- ▶ Les deux annonces contiennent beaucoup d'information, mais difficile de trancher
- ▶ Pourquoi ? Parce qu'en réalité ce qui compte ici, ce n'est pas le contenu ou le sens de l'information, mais plutôt la surprise que cela va créer chez le destinataire. Dans les deux cas, comme les deux annonces sont très peu probables, elles nous apportent beaucoup d'information.
- ▶ Shannon est parti de la même supposition : la quantité d'information d'un message ne doit pas dépendre du sens ou de la sémantique de ce message, mais uniquement de la probabilité d'apparition du message.

## Tentative de définition

- ▶ Notion d'information et de quantification de l'information  
→ On va être capable de définir une valeur numérique concrète quantifiant l'information contenue dans un message donné
- ▶ Vision probabiliste de l'information  
→ L'information contenue dans un message ne dépend pas du sens sémantique du message, mais uniquement de sa probabilité d'apparition
- ▶ Lien avec le codage  
→ Connaissant les différents messages à envoyer et leur probabilité d'apparition, on va être capable de leur associer un code *intelligent* permettant de compresser et de transmettre au mieux les données