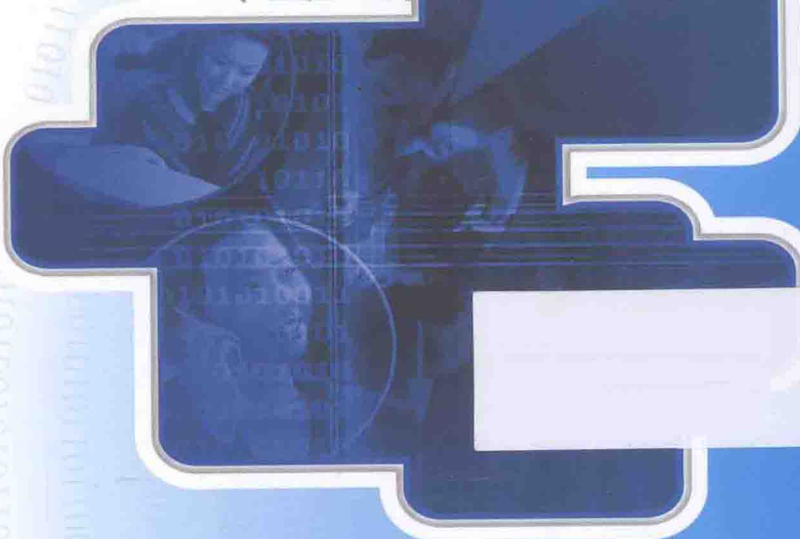
21世纪全国本科院校电气信息类创新型应用人才培养规划教材

信息与通信专业英语

刘小佳 主编



提供
电子课件
www.pup6.cn

北京大学出版社
PEKING UNIVERSITY PRESS

21世纪全国本科院校电气信息类创新型应用人才培养规划教材

信息与通信专业英语

主　编　刘小佳

北京大学出版社
PEKING UNIVERSITY PRESS

内 容 简 介

本书由课文、阅读材料和摘要阅读组成，内容分别选自各领域经典英文教材、国际顶级期刊及会议的论文摘要，紧扣信息、通信与计算三大主题，内容全面详实。

本书可作为信息与通信工程、信息安全、电子科学与技术、电子信息工程、计算机科学与技术和网络工程等专业本科生、研究生“专业英语”课程的教材，也可供相关专业技术人员学习和参考。

图书在版编目(CIP)数据

信息与通信专业英语/刘小佳主编．—北京：北京大学出版社，2015．4
（21 世纪全国本科院校电气信息类创新型应用人才培养规划教材）
ISBN 978－7－301－25506－3

Ⅰ．①信… Ⅱ．①刘… Ⅲ．①信息技术—英语—高等学校—教材②通信工程—英语—高等学校—教材 Ⅳ．①H31

中国版本图书馆 CIP 数据核字（2015）第 031991 号

书　　名	信息与通信专业英语
著作责任者	刘小佳　主编
责 任 编 辑	程志强
标 准 书 号	ISBN 978－7－301－25506－3
出 版 发 行	北京大学出版社
地　　址	北京市海淀区成府路 205 号　100871
网　　址	http://www.pup.cn　新浪微博：@北京大学出版社
电 子 信 箱	pup_6@163.com
电　　话	邮购部 62752015　发行部 62750672　编辑部 62750667
印 刷 者	北京富生印刷厂
经 销 者	新华书店
	787 毫米×1092 毫米　16 开本　13.75 印张　312 千字
	2015 年 4 月第 1 版　2015 年 4 月第 1 次印刷
定　　价	29.00 元

前　　言

多年来，专业英语教学一直想传达本科教育的国际化视野之理念，但收效甚微。许多专业英语教材尽管选择了非常地道的英文原文，但其内容止于“科普”，未能达到“专业”。事实上，文史哲中所推崇的原典阅读是一个极好的思路，本书正是在此基础上展开了有益的尝试。

首先，要能阅读较为专业的英语文本。作者从电子信息类与通信类专业领域经典英文教材入手，精选若干原文，模拟学生学习该专业课程原文教材的场景，以此提高专业英语的针对性。显然，这种直接阅读大师原典的做法不但能提高使用专业英语的能力，还能在专业方面打开知识的大门，为进一步阅读经典教材全文打下基础。

其次，专业英语写作也是很重要的内容。作者在每个领域选出了若干国际顶级期刊与会议的论文摘要，针对国内学生摘要写作较差的特点进行强化训练，以此培养他们科技论文写作的能力，并以摘要为起点抛砖引玉，令有能力的学生进行完整的英语科技论文写作。

在内容选取上，本书充分考虑了通信工程、信息安全、电子科学与技术、电子信息工程、计算机科学与技术和网络工程等专业的特色，紧扣信息、通信与计算三大主题，突出这三个领域的交叉与融合，不但为更多专业的学生提供了学习的内容，还注重培养各专业对其相关领域的关注了解，从而更好地提高学生的综合实力。

本书共分为16章，每章由一篇课文、一篇阅读材料和三篇摘要阅读组成。读者在学习完课文之后，应精心理解阅读材料。对于学有余力的学生，还应该寻根溯源去广泛阅读对应的国外教材，这样才能达到熟练掌握该领域英语的功效。为了帮助读者学习，作者对课文中的词汇给出了在该领域的准确释义，并精选课文中的阅读难点予以注释。需要指出的是，读者应仔细研读摘要的写法，不但能将英文翻译为中文，还能将已翻译的中文回译成英文，多年的教学实践证明，这种对比参照的双向翻译能收到良好的效果。

感谢西安邮电大学外国语学院袁小陆教授仔细审阅了全部书稿，并提出了许多中肯的建议！感谢西安邮电大学外国语学院诸多同事在教材编写中的鼎力相助！本书摘选诸多国外经典教材的英文文本，并遴选多篇顶级期刊会议的论文摘要，在此一并对原作者致谢！

由于编者学识所限，书中疏漏之处在所难免，恳请读者不吝赐教！

编　者

2014年12月

目　　录

UNIT 1

INFORMATION THEORY

Text: Introduction to Information Theory

Information theory answers two fundamental questions in communication theory: What is the ultimate data compression (answer: the entropy H), and what is the ultimate transmission rate of communication (answer: the channel capacity C). For this reason some consider information theory to be a subset of communication theory. We argue that it is much more. Indeed, it has fundamental contributions to make in statistical physics (thermodynamics), computer science (Kolmogorov complexity or algorithmic complexity), statistical inference (Occam's Razor: "The simplest explanation is best"), and to probability and statistics (error exponents for optimal hypothesis testing and estimation).

This chapter goes backward and forward through information theory and its naturally related ideas. Information theory intersects physics (statistical mechanics), mathematics (probability theory), electrical engineering (communication theory), and computer science (algorithmic complexity). We now describe the areas of intersection in greater detail.

Electrical Engineering (Communication Theory). In the early 1940s it was thought to be impossible to send information at a positive rate with negligible probability of error. Shannon surprised the communication theory community by proving that the probability of error could be made nearly zero for all communication rates below channel capacity.

The capacity can be computed simply from the noise characteristics of the channel[1]. Shannon further argued that random processes such as music and speech have an irreducible complexity below which the signal cannot be compressed. This he named the entropy, in deference to the parallel use of this word in thermodynamics, and argued that if the entropy of the source is less than the capacity of the channel, asymptotically error-free communication can be achieved.

Information theory today represents the extreme points of the set of all possible communication schemes. The data compression minimum $I(X;\hat{X})$ lies at one extreme of the set of communication ideas. All data compression schemes require description rates at least equal to this minimum. At the other extreme is the data transmission maximum $I(X;Y)$, known as the channel capacity. Thus, all modulation schemes and data compression schemes lie between these limits.

Information theory also suggests means of achieving these ultimate limits of communication. However, these theoretically optimal communication schemes, beautiful as they are, may turn out to be computationally impractical. It is only because of the computational feasibility of simple modulation and demodulation schemes that we use them rather than the random coding and nearest-neighbor decoding rule suggested byShannon's proof of the channel capacity theorem. Progress in integrated circuits and code design has enabled us to reap some of the gains suggested by Shannon's theory. Computational practicality was finally achieved by the advent of turbo codes. A good example of an application of the ideas of information theory is the use of error-correcting codes on compact discs and DVDs.

Recent work on the communication aspects of information theory has concentrated on network information theory: the theory of the simultaneous rates of communication from many senders to many receivers in the presence of interference and noise. Some of the trade-offs of rates between senders and receivers are unexpected, and all have a certain mathematical simplicity. A unifying theory, however, remains to be found.

Computer Science (Kolmogorov Complexity). Kolmogorov, Chaitin, and Solomonoff put forth the idea that the complexity of a string of data can be defined by the length of the shortest binary computer program for computing the string. Thus, the complexity is the minimal description length. This definition of complexity turns out to be universal, that is, computer independent, and is of fundamental importance[2]. Thus, Kolmogorov complexity lays the foundation for the theory of descriptive complexity. Gratifyingly, the Kolmogorov complexity K is approximately equal to the Shannon entropy H if the sequence is drawn at random from a distribution that has entropy H. So the tie-in between information theory and Kolmogorov complexity is perfect. Indeed, we consider Kolmogorov complexity to be more fundamental than Shannon entropy. It is the ultimate data compression and leads to a logically consistent procedure for inference.

There is a pleasing complementary relationship between algorithmic complexity and computational complexity. One can think about computational complexity (time complexity) and Kolmogorov complexity (program length or descriptive complexity) as two axes corresponding to program running time and program length. Kolmogorov complexity focuses on minimizing along the second axis, and computational complexity focuses on minimizing along the first axis. Little work has been done on the simultaneous minimization of the two.

Physics(Thermodynamics). Statistical mechanics is the birthplace of entropy and the second law of thermodynamics. Entropy always increases. Among other things, the second law allows one to dismiss any claims to perpetual motion machines.

Mathematics(Probability Theory and Statistics). The fundamental quantities of information theory—entropy, relative entropy, and mutual information—are defined as functionals of probability distributions[3]. In turn, they characterize the behavior of long sequences of random variables and allow us to estimate the probabilities of rare events (large deviation theory) and to find the best error exponent in hypothesis tests.

Philosophy of Science (Occam's Razor). William of Occam said "Causes shall not be multiplied beyond necessity," or to paraphrase it, "The simplest explanation is best." [4] Solomonoff and Chaitin argued persuasively that one gets a universally good prediction procedure if one takes a weighted combination of all programs that explain the data and observes what they print next. Moreover, this inference will work in many problems not handled by statistics. For example, this procedure will eventually predict the subsequent digits of π. When this procedure is applied to coin flips that come up heads with probability 0.7, this too will be inferred. When applied to the stock market, the procedure should essentially find all the "laws" of the stock market and extrapolate them optimally. In principle, such a procedure would have found Newton's laws of physics. Of course, such inference is highly impractical, because weeding out all computer programs that fail to generate existing data will take impossibly long. We would predict what happens tomorrow a hundred years from now.

Economics (Investment). Repeated investment in a stationary stock market results in an exponential growth of wealth. The growth rate of the wealth is a dual of the entropy rate of the stock market. The parallels between the theory of optimal investment in the stock market and information theory are striking. We develop the theory of investment to explore this duality.

Computation vs. Communication. As we build larger computers out of smaller components, we encounter both a computation limit and a communication limit. Computation is communication limited and communication is computation limited. These become intertwined, and thus all of the developments in communication theory via information theory should have a direct impact on the theory of computation.

New Words and Phrases

transmission *n.* 传输
subset *n.* 子集
thermodynamics *n.* 热力学
algorithmic *adj.* 算法的
statistical inference 统计推断
exponent *n.* 指数
optimal *adj.* 最佳的，最优的
hypothesis *n.* 假设
intersect *vi.* 交叉，相交
complexity *n.* 复杂度
universal *adj.* 通用的，广泛的，普遍的，宇宙的
intersection *n.* 交集
mutual *adj.* 相互的，共同的
negligible *adj.* 微小的；可忽略的
irreducible *adj.* 不能分解的
entropy *n.* 熵
probability *n.* 概率
deference *n.* 顺从；尊重
asymptotically *adv.* 渐近地
modulation *n.* 调制
feasibility *n.* 可行性
integrated circuit 集成电路
reap *vt.* & *vi.* 收获
turbo code turbo 编码
simultaneous *adj.* 同时的，同步的
trade-off *n.* 权衡
unifying *vt.* 使统一
binary *adj.* 二进制的，二元的
tie-in *n.* 接头
compression *n.* 压缩
complementary *adj.* 互补的，补充的
axis *n.* 轴，轴线
perpetual motion 永恒运动
deviation *n.* 偏差；偏向
subsequent *adj.* 后来的，随后的
extrapolate *vt.* & *vi.* 推断，推算
optimally *adj.* 最佳
weed out 淘汰，剔除
exponential *n.* 指数
dual *n.* 对偶
adj. 指数的
intertwine *vi.* & *vt.* 交织，纠缠
theorem *n.* 定理

Notes

1. The capacity can be computed simply from the noise characteristics of the channel.
（信道）容量只用信道的噪声特征即可计算。

2. Thus, the complexity is the minimal description length. This definition of complexity turns out to be universal, that is, computer independent, and is of fundamental importance.

因此，复杂度是最小描述长度。这个复杂度定义是通用的，也就是说，它与具体计算机无关并且有着根本的重要性。

3. The fundamental quantities of information theory—entropy, relative entropy, and mutual information—are defined as functionals of probability distributions.

信息论的基本量——熵、相对熵、互信息——被定义为与概率分布有关的函数。

4. (Occam's Razor) "Causes shall not be multiplied beyond necessity," or to paraphrase it, "The simplest explanation is best."

"如无必要，勿增缘由"，也就是"简单的解释就是最好的"。

Occam's Razor 奥卡姆剃刀原理：如果你有两个原理，它们都能解释观测到的事实，那么你应该使用简单的那个，直到发现更多的证据。对于现象最简单的解释往往比复杂的解释更正确。如果你有两个类似的解决方案，选择最简单的。需要最少假设的解释最有可能是正确的。

Exercises

Ⅰ. Please translate the following words and phrases into Chinese.

1. error-correcting
2. probability theory
3. algorithmic complexity
4. large deviation theory
5. random processes
6. modulation schemes
7. statistical inference
8. negligible probability

Ⅱ. Fill in the blanks with the missing word(s) from the table below.

theorem	compression	elements	length
capacity	entropy	random	probability
codes	parallels	transmission	mutual
defined	block	output	exponent
continuous	duality	redundancy	constructing
satisfy	ratio	capacity	concept

1. The relative ________ D arises as the ________ in the probability of error in a hypothesis test between two distributions. It is a natural measure of distance between distributions.

2. There are a number of ________ between information theory and the theory of investment in a stock market. A stock market is ________ by a random vector X whose ________are nonnegative numbers equal to the ________ of the price of a stock at the end of a day to the price at the beginning of the day.

3. Entropy is the uncertainty of a single ________ variable. We can define conditional entropy $H(X \mid Y)$, which is the entropy of a random variable conditional on the knowledge of another random variable.

4. We now define the (information) ________ of the channel as the maximum of the ________ information between the input and ________ over all distributions on the input that ________ the power constraint.

5. We now introduce the concept of differential entropy, which is the entropy of a ________ random variable. Differential entropy is also related to the shortest description ________ and is similar in many ways to the entropy of a discrete random variable. But there are some important differences, and there is need for some care in using the ________.

6. The channel coding ________ promises the existence of block codes that will allow us to transmit information at rates below ________ with an arbitrarily small ________ of error if the block length is large enough.

7. Although the theorem shows that there exist good ________ with arbitrarily small probability of error for long ________ lengths, it does not provide a way of ________ the best codes.

8. There is a ________ between the problems of data ________ and data transmission. During compression, we remove all the ________ in the data to form the most compressed version possible, whereas during data ________, we add redundancy in a controlled fashion to combat errors in the channel.

Ⅲ. Translate the following paragraphs into Chinese.

1. At first sight, information theory and gambling seem to be unrelated. But as we shall see, there is strong duality between the growth rate of investment in a horse race and the entropy rate of the horse race. Indeed, the sum of the growth rate and the entropy rate is a constant. In the process of proving this, we shall argue that the financial value of side information is equal to the mutual information between the horse race and the side information. The horse race is a special case of investment in the stock market. We also show how to use a pair of identical gamblers to compress a sequence of random variables by an amount equal to the growth rate of wealth on that sequence. Finally, we use these gambling techniques to estimate the entropy rate of English.

2. This enables us to divide the set of all sequences into two sets, the typical set, where the sample entropy is close to the true entropy, and the nontypical set, which contains the other sequences. Most of our attention will be on the typical sequences. Any property that is proved for the typical sequences will then be true with high probability and will determine the average behavior of a large sample.

Reading: Channel Capacity

What do we mean when we say that A communicates with B? We mean that the physical acts of A have induced a desired physical state in B. This transfer of information is a physical process and therefore is subject to the uncontrollable ambient noise and imperfections of the physical signaling process itself. The communication is successful if the receiver B and the transmitter A agree on what was sent.

We find the maximum number of distinguishable signals for n uses of a communication channel. This number grows exponentially with n, and the exponent is known as the channel capacity. The characterization of the channel capacity (the logarithm of the number of distinguishable signals) as the maximum mutual information is the central and most famous success of information theory.

The mathematical analog of a physical signaling system is shown in Figure 1. 1. Source symbols from some finite alphabet are mapped into some sequence of channel symbols, which then produces the output sequence of the channel. The output sequence is random but has a distribution that depends on the input sequence. From the output sequence, we attempt to recover the transmitted message.

Each of the possible input sequences induces a probability distribution on the output sequences. Since two different input sequences may give rise to the same output sequence, the inputs are confusable. In the next few sections, we show that we can choose a "nonconfusable" subset of input sequences so that with high probability there is only one highly likely input that could have caused the particular output. We can then reconstruct the input sequences at the output with a negligible probability of error. By mapping the source into the appropriate "widely spaced" input sequences to the channel, we can transmit a message with very low probability of error and reconstruct the source message at the output. The maximum rate at which this can be done is called the capacity of the channel.

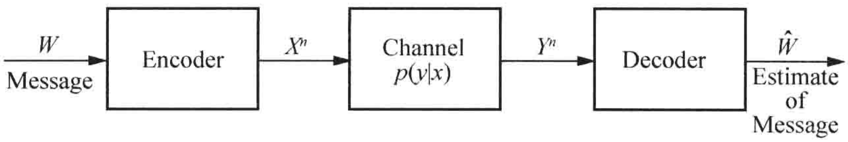


Figure 1. 1 Communication System

Definition We define *a discrete channel* to be a system consisting of an input alphabet X and output alphabet Y and a probability transition matrix $p(y \mid x)$ that expresses the probability of observing the output symbol y given that we send the symbol x. The channel is said to be *memoryless* if the probability distribution of the output depends only on the input at that time and is conditionally independent of previous channel inputs or outputs.

Definition We define the "*information*" *channel capacity* of a discrete memoryless channel as where the maximum is taken over all possible input distributions$p(x)$.

$$C = \max_{p(x)} I(X;Y) \tag{1-1}$$

We shall soon give an operational definition of channel capacity as the highest rate in bits per channel use at which information can be sent with arbitrarily low probability of error. Shannon's second theorem establishes that the information channel capacity is equal to the operational channel capacity. Thus, we drop the word information in most discussions of channel capacity.

There is a duality between the problems of data compression and data transmission. During compression, we remove all the redundancy in the data to form the most compressed version possible, whereas during data transmission, we add redundancy in a controlled fashion to combat errors in the channel. Later we show that a general communication system can be broken into two parts and that the problems of data compression and data transmission can be considered separately.

New Words and Phrases

ambient *adj.* 背景的
imperfection *n.* 缺陷
exponentially *adv.* 成指数地
exponent *n.* 指数，幂
analog *n.* 模拟
negligible *adj.* 可忽略不计的
discrete *adj.* 离散，离散的
arbitrarily *adv.* 任意地
duality *n.* 对偶，二元性
compression *n.* 压缩
redundancy *n.* 冗余，冗余度

Exercises

Ⅰ. Answer the following questions.

1. What is the discrete channel?
2. How to calculate the channel capacity?
3. What is theredundancy?

Ⅱ. Translate the following sentences into Chinese.

1. This transfer of information is a physical process and therefore is subject to the uncontrollable ambient noise and imperfections of the physical signaling process itself.

____

____

2. Source symbols from some finite alphabet are mapped into some sequence of channel symbols, which then produces the output sequence of the channel.

____

____

3. By mapping the source into the appropriate "widely spaced" input sequences to the channel, we can transmit a message with very low probability of error and reconstruct the source message at the output.

____

____

4. During compression, we remove all the redundancy in the data to form the most compressed version possible, whereas during data transmission, we add redundancy in a controlled fashion to combat errors in the channel.

____

____

____

Abstract Reading

On the Capacity of the Two-user Gaussian Causal Cognitive Interference Channel

This paper considers the two-user Gaussian causal cognitive interference channel (GCCIC), which consists of two source-destination pairs that share the same channel and where one full-duplex cognitive source can causally learn the message of the primary source through a noisy link. The GCCIC is an interference channel with unilateral source cooperation that better models practical cognitive radio networks than the commonly used model which assumes that one source has perfect noncausal knowledge of the other source's message. First, the sum-capacity of the symmetric GCCIC is determined to within a constant gap. Then, the insights gained from the study of the symmetric GCCIC are extended to more general cases. In particular, the whole capacity region of the Gaussian Z-channel, i. e. , when there is no interference from the primary user, and of the Gaussian S-channel, i. e. , when there is no interference from the secondary user, are both characterized to within 2 bits. The fully connected general, i. e. , no-symmetric, GCCIC is also considered and its capacity region is characterized to within 2 bits when, roughly speaking, the interference is not weak at both receivers. The parameter regimes where the GCCIC is equivalent, in terms of generalized degrees-of-freedom, to the noncooperative interference channel (i. e. , unilateral causal cooperation is not useful), to the non-causal cognitive interference channel (i. e. , causal cooperation attains the ultimate limit of cognitive radio technology), and to bilateral source cooperation are identified. These comparisons shed light into the parameter regimes and network topologies that in practice might provide an unbounded throughput gain compared to currently available (non cognitive) technologies.

Channel Coding and Lossy Source Coding Using a Generator of Constrained Random Numbers

Stochastic encoders for channel coding and lossy source coding are introduced with a