CHAPTER-3
CONCEPTUAL TRANSFER
FROM SEMANTIC SPACE
Chapter 3

CONCEPTUAL TRANSFER FROM SEMANTIC SPACE

3.0 Introduction:
This chapter is a try to through light that the Translation is more than just a linguistic activity; translation is one of the main ways in which intercultural relationships are formed and transformed. The study of translation should thus involve far more than merely defining and testing linguistic equivalents. It should ask what relation translation has to the texts that move between cultures; it should have ideas about why texts move and how translated texts can represent such movement; and it should be able to inquire into the ethics of intercultural relations and how translators should respond them.

In short, by relating the work of translators to the problematics of intercultural transfer, translation studies should take its rightful interdisciplinary place among the social sciences, arts and sciences.

But what kind of conceptual geometry might make this development possible? Refusing simple answers, this chapter sees the relation between translation and transfer as a complex phenomenon that must be described on both the semiotic and material levels.

Various connected approaches then conceptualize this relationship as being causal, economic, discursive, quantitative, political, historical, ethical and epistemological and indeed translational. Individual chapters address each of these aspects, placing particular emphasis on phenomena that are mostly ignored by contemporary theories.
The result is a dense but highly suggestive and hopefully stimulating vision of translation studies.

The principle, that the translation should produce the same effect on the reader as the original text. Taking these issues and more into the consideration this chapter has been divided into following sections.

3.1 Transferability
3.2 Equivalence
3.3 Dictionary based translation
3.4 Selection of ‘Keywords’
3.5 Disambiguation
3.6 Word based Disambiguation
3.7 The Mutual Information Leading to Disambiguation
3.8 Conclusion

3.1 Transferability

This topic considers the issue of transferability, a well-known concept in the translation of literature from source language L1 to the target language L2. Transferability means exchangeability; fungibility; interchangeability; interchangeableness. The quality of being, capable of exchange or interchange

Transferability thus has logical priority as a necessary precondition for the general practice of translating. If nothing has moved or is going to move from L1 to L2, then there is no reason to translate from the culture of L1 into the culture of L2. If someone is translating or has translated, then something has moved or is meant to move.
It is easy enough to say that a ‘text’ is what is transferred. But exactly what is a text? How do we recognize one when we see it? And how do we recognize the fact that it has been or is being transferred?

In order to answer these questions properly, we need a clear idea of the kind of transfer relevant to translation. How can we identify whatever it is that moves in the process of transfer?

This simple principle underlies the rather more interesting proposition that translation studies should accord more priority to the movements of objects than to those of subjects.

Although theories of translation rarely talk about transfer as such, they do tend to make assumptions about what can ideally be taken from one culture to another.

For some, there is no real movement, since the one mark always approximates the same pre-existing “meaning” or “concept”: if the handprint meant “hand” when it was made and it means “hand” when received by X and Y, how could one say that anything has moved?

Universalist semantics wants us to believe that everything was always already there. In this way, blindness to transfer does away with the basic reasons for translation. For other theories, however, there is real movement in the sense that the mark functions as “information”, “signification”, a “message” or even “enlightenment”, bringing new meaning to the particular receivers. This approach can at least explain why there should have been an act of transfer and thus the possibility of translation. But does it really matter what the mark might have meant before it reached these new receivers?
Now what is text? Texts—including only oral texts—ever exist without the materiality of a support? Does their status as an object of knowledge ever not presuppose a level of substance? The kind of transfer that goes from rock to paper or film requires that the supports come into material contact or proximity with the inscribed form.

When, as in the case of the rock-mark or pictographs such as Ajanta Alora caves etc, such simple transfer is across time instead of space, the contact between form and support is continuous. But the principle of necessary materiality is the same in both cases. It is impossible to find a text devoid of a support, be it rock, electrons, genes, sound waves, or whatever else is able to go from one point to another.

The nature of the support can change—from rock to paper, from paper to voice, but at no point is the text liberated from the materiality of things that move. To imagine otherwise is to pretend that texts fall from the sky and exist forever. For anyone is not just concerned with the hand-shape or with the text as form. They find a text whose materiality indicates it has come from another time.

Flaking ochre and weathered rock must say more about the handprint’s age than does the simple absence of its producer. Reception is concerned with a text which is, hand-shape and rock, form and material, since both these aspects are necessary if the receivers are to conclude that the absent producer is now long dead.

This textual materiality allows receiver to attach importance to the physical dimensions of the text—the shapes are of a certain size, then use comparison to attribute meaning to that size.

Transfer thus enables a process of interpretation, a comparison, a figuration of the absent producer, a potential utterance and a complex contextual meaning as an artificial alternative to natural reproduction. For person, as for most of Stow’s
heroes, writing will sublimate sexuality as transfer of the self. More importantly, no one need insist that ancient piccaninny had any such meaning in mind.

Transfer in this case enables a process of interpretation which borders translation. Simple transfer might thus be enough for some form of knowledge to be produced through intercultural communication.

But is there any translation here? Is there any strictly translated text? If the second, interpretative hand had not been evaluated as significantly larger than the transferred text, if it had not been conceptually attached to the interpreting subjectivity known as receiver, then it might have been possible to consider it as a translation.

Or again, perhaps one could consider receiver as a potential translator, the translated text then being his/ her phrase “It’s a child’s”. But the deictic “it’s” separates the object transferred from the subject translating, in the same way as sheer size separates the textual hand from anybody’s interpreting hand.

People cannot be seen as translating the pictures; they simply comment on it as an object external to their own time and place. There is a difference between translating a text and just talking about it or producing a similar text.

Translations are quite difficult to achieve; they are very particular kinds of communicative artifacts. As we have seen, not all acts of transfer need give rise to complete acts of translation. And as we shall see in the following topics, translations moreover require fulfillment of a series of specific conditions which go well beyond transfer, including a certain kind of belief on the part of the person receiving the translated text. Exactly what is transferred? For the purposes of translation studies, the privileged object of transfer is the text, independently of
whatever meaning, information, message or signification might have been attributed to that text prior to transfer.

But the text must be recognized as inseparable from material support, since it is only through materiality that its transfer can become significant.

The principle of meaningful materiality involves theoretical consequences well beyond our immediate concerns. It is possible, for example, that coherence and cohesion presuppose a continuity of material support both before and during reception, even when this continuity is not realized because of broken or ruptured transmission.

It is conceivable that fanfares of inter-textuality should be limited by quite reasonable criteria of historical contiguity: if there is to have been some kind of transfer from one text to another, then the two texts concerned must at some time have shared the same locus.

But the important point for our present purposes is that the necessary materiality of texts condemns them to displacement. Indeed, not only are texts always available for transfer, they are by definition unable to avoid being transferred, through time if not always through space.

3.1.1 Transferability can be intra-lingual or inter-lingual

It is often assumed that the kind of transfer most pertinent to translation is that which takes place exclusively between different languages. This restriction of the field assumes a radical division between inter-lingual and intra-lingual transfer. Unfortunately there is no such division, simply because there are no natural frontiers between languages. The kinds of translation that can take place between idiolects, sociolects and dialects are essentially no different from those between more radically distanced language systems.
Consider, for example, the various transformations necessary to rewrite in the English of Queen Elizabeth II a text from American English, working-class Liverpudlian, Shakespeare’s English, Chaucer’s English, the French of François Mitterrand and Japanese. Deputy Nazir Ahmad’s Urdu that is 19th century Urdu and today’s Urdu.

Although one would expect to encounter a need for increasing transformations with increasing cultural distance, there is no strict cut-off point at which wholly intralingual rewriting can be said to have become wholly interlingual.

Those who travel on foot or have read the diachronic part of Saussure know that there are no natural frontiers between languages.

Since “language A” and “language B” are insufficient descriptions of the two places minimally involved in translation, some alternative vocabulary must be sought. A Chomskyan “ideal speaker-listener in a completely homogeneous language community” (1965, 3) would clearly be inadequate for much the same reasons as “language A”: since there are many more languages in the world than countries to house them, the fact of bilingual and polyglot communities must be recognized and incorporated into any global approach to translation.

Similarly, since numerous languages are spoken in more than one community, it must be admitted that texts can be transferred from one community to another and yet not require translation because the original language of the text is able to seek out its appropriate receivers.

Neither “language” nor “community” is sufficient criteria for the description of the kinds of places minimally involved in translation. A certain retreat to the bunker is necessary, in this case to the suitably vague term “culture”.

That is, the kind of transfer we consider pertinent to translation is that which takes place between different cultures.

But what then is a culture? How might one define the points where one culture stops and another begins? The borders are no easier to draw than those between languages or communities. One could perhaps turn to a geometry of fuzzy sets or maybe even deny the possibility of real contact altogether, but neither mathematics nor ideological relativism are able to elucidate the specific importance of translation as an active relation between cultures.

Although questions like the definition of a culture are commonly thought to lie beyond the scope of translation theory, their solution could become one of translation studies’ main contributions to the social sciences.

Instead of looking for differentiated or distilled cultural essences, it could be fruitful to look at translations themselves in order to see what they have to say about cultural frontiers. It is enough to define the limits of a culture as the points where transferred texts have had to be (intralingually or interlingually) translated. That is, if a text can adequately be transferred without translation, there is cultural continuity. And if a text has been translated, it represents distance between at least two cultures.

In this way, translation studies avoid having to link up all the points of contiguity in the way that political frontiers do. After all, there is no obvious reason why point’s of contact and exchange between cultures should form continuous lines. Culture is not geo-politics. Transfer and translation concern situations of contact and exchange, not lineal separations.

According to the solidarity of these definitions, specifically intercultural transfer is a precondition for general translation, and translation itself therefore logically
indicates both the existence of intercultural transfer and the points separating the cultures concerned.

Instead of using preconceptions about cultures in order to form preconceptions about translations, it is thus possible to use facts about translations in order to locate contacts and differences between cultures. Indeed, to do so could be conceptually elegant.

3.1.2 Translation can be approached from Transfer

These few comments on the nature of transfer provide us with two basic ways of approaching its relation with translation. On the one hand, translation is partly knowable through the analysis of texts which have been translated (or, more ambiguously, through the past-participle form “translated text” or TT, which, from the perspective of the translating translator, can also be read as “target language text” TLT).

On the other, to know why and how any particular translational operation was or should be carried out, we have to look at the factors involved in the transfer from a distanced or even imaginary source text (ST) or source language text (SLT) to the place of a manifest TLT.

We have to ask what came from where and for what reason; and where, why and to whom the translated text is to go.

Two complementary approaches are thus available from the outset: one is textual (translation as representation), the other is extra-textual (translation as response).

Almost everyone interrogates translation from the first of these perspectives, making vast use of semiotic science and diverse cultural convictions but in fact basing their observations on no more than translated texts as representations.
However, the second set of questions, deceptively simple, can often subvert the conclusions thus reached.

3.1.3 Transferability against Belonging

One of the recurrent shortcomings in general approaches to translation is that we are all constantly prepared to say how translation works but we rarely stop to ask what it might be working against.

If, according to a not uncommon Manichean vision, translation were all goodness and enlightenment, what would then be the badness and ignorance it should oppose? And even if translation were found to be less than ideal, where should we look to find the values it lacks? In short, what is the opposite of translation?

This is a facile evasion of the issue since, as we have seen, the decision not to translate is still a translational decision; it puts no illogical spanner into the works of translation as a process; its minor elitism is entirely describable in terms of the basic principles of translational quantities.

Transliteration might be a degree of apparent non-translation, but it is certainly not anti-translational.

A far more formidable opposite lies not within translation as such, but in the negation of transfer, in denial of the necessary precondition for any translational situation. Maximum anti-translationality is surely to be found somewhere in non-transfer, in the state of texts whose movements are so restricted that they almost never become candidates for properly translational treatment. Translators are most successfully thwarted by the material they cannot get their hands on; translation can fail because some aspects of texts are difficult to transfer situation and sometimes impossible to convey to the situation of translational reception.
The fundamental task of translators must thus be to work against the bonds restricting the movements of texts; they must attempt to attain some kind of maximum trans-cultural mobility towards a specific receiver. In short, translation works against constraints on transfer.

Why should there be constraints on the transfer of texts? Quite simply because reading “transfer” in its juridical sense, certain texts are felt to belong to certain people or to certain situations, making the movement of those texts a question of bestowing some kind of ownership or right to full textual meaningfulness. Seen in this light, constraints on transfer can be thought of as bonds of belonging.

3.1.4 There are no solo performances in Transferability

Explicit performatives are no longer performative once translated. An utterance like “I declare games are open” may well be translated as ‘Main khelon ke aGhaas ka aelaan karta hun’ in Urdu, but only the first, non-translational version whatever its tongue-can actually open the games. This is a simple example of a text that belongs in one place and not in another, denying full meaningfulness in the translational situation.

We could say that translations of performatives are necessarily constative, or, to adopt the terminology of Christiane Nord (1988), the instrumental status of the non-translated text becomes documental when translated.

In a sense, the instrumental text—the performing performative, the text as action—is owned by the people who originally open the meeting, and cannot be owned in the same way by those who receive only a report of it.

In terms of translatability, the instrumentality or effective performance of performatives obviously has nothing to do with language choices. It does not
matter which language is used to open the meeting, since only the first language to be used will be the one to belong. What resists transfer is not language, but the text’s situational ability to become part of an action. We are concerned only with this ability to resist transfer.

3.1.5 Distance can break performance in Transferability

Since the discursive second person is defined simply as a potentially participative “non-I”, there is in principle no distance between the two positions concerned. If the performance is to be fully performative, both “I” and “you” must be mutually present.

Unfortunately, beyond the world of discursive positions, there is always time and space between the entities filling the “I” and the “you”. This extra-textual distance interests us because it might be able to determine thresholds beyond which certain texts cease to be properly instrumental. The isolated question “What is the time?” a key example for Bakhtin’s dialogic principle (1930, 164ff.) is not only discursively unique in terms of the “now” of each performance, but also implies the presence of a receiver located very close to the I-here-now.

That is, unavoidable factors like the distance involved in the passage to a real receiver, the processing of the message, the formulation of a reply and the sending and reception of the reply should, according to discursive logic, not involve any time at all, since the “now” of the question sent should correspond to the “now” of the reply received. And yet it is obvious that the entire double transfer process must take time and the two “now’s cannot be the same.

The discursive logic is at odds with the fact of material distance. For how long can discursive logic win out? For how long can participants pretend that there is no distance between them, that the “now” of the question is referred to by the “now”
of the reply? within a few seconds perhaps, if the speaker’s requirements do not include extreme exactitude. But a minute would create an anomalous and slightly absurd situation, and it is difficult to imagine a situation in which the isolated question could be written. The utterance “What is the time?” can only be transferred away from its “I-here-now” within certain pragmatic limits; it has a certain elasticity, a capacity to stretch out and remain valid as meaningful communication. But beyond a certain point, this elasticity will be broken and communication will be defeated by distance. The utterance will reveal its limited transferability. It is possible that the relative transferability of a text can be analyzed simply by isolating and assessing all those elements that refer to or imply an I-here-now calling for participation by a second person.

3.1.6 Transferability has second-person Thresholds

The analysis of language as action has its limits. After all, not all texts require the same kind of immediate response as “What is the time?” If I write and send a letter in search of a reply, the second person should presumably be located beyond the range of my voice but not beyond the range of my expected lifetime or the value of the stamps put on the envelope. In this case, it would make no sense to ask for the time.

Participative second persons can be sought on at least three levels, associated with three different degrees of transfer:

- **Shared transfer**: For some texts, the implied receiver calls for an actual receiver who is situationally present so that a reply can reach the sender and discursively share the same I-here-now. That is, the reply should ideally be in the same present tense (“The time is...”, or “Do you mean that...?”). Transfer beyond this restricted
and usually oral situation will leave the text decidedly out of place. (Example: “What is the time?”)

- **Referential transfer**: Where the physical receiver called for by the implied receiver need not share the same I-here-now but should be able to refer to at least one of these coordinates, transfer may well extend beyond short distance situational constraints. But if some reply is required—if the text is not indifferent to the location of the second person, factors of age, space and time impose practical constraints on transferability. (Example: Ghalib ke Khutoot)

- **Indefinite transfer**: Some texts need neither reply nor reference to an I-here-now; they can in principle be transferred indefinitely, within the constraints of world and time. (Example: Ghalib ke Khutoo)

The thresholds between these three modes are obviously imposed on a practical continuum that goes from short-distance to mid-distance to long-distance transfer, regulating the degrees to which increased distance may break initial elasticity.

It might prove possible to relate these thresholds to particular text-types, perhaps along the lines of the categories Reiss (1971) adopts from Bühler, originally based on the three linguistic persons. But such correlations are more apparent than real, since transfer-based analysis concerns the position and role of the second person within an inclusive “we”; it is in principle indifferent to the relative dominance of first or third persons or to relative focus on form, content or effect.

In fact, the problem of text-types could turn out to be something of a red herring in this context. Although it is true that the limited elasticity of certain texts restricts their ability to remain properly instrumental after various degrees of transfer, there is no fatal determinism at stake: a text can be transferred beyond its initial elasticity and assume a new performative status; or the same discourse can be
rewritten, with the same pronominal structure but with more elaborate coordinates, in order to heighten its transferability. That is, texts can be transformed in order to pass over one threshold or another, and the modes of rewriting need not affect text-type models based on criteria other than those of transferability.

3.1.7 Textual worlds increase Transferability

The possibility of breaking bonds of belonging and yet maintaining instrumental status clearly has something to do with referentiality and the existence of codes shared by separated sending and receiving positions. But it would be inadequate to describe transferability as a quality of the referents or the codes themselves. Its conditions are not strictly semantic; its investigation does not present problems as to whether or not we really understand the intentions that originally gave rise to the utterances.

Transferability has much more to do with how well the text itself can function as a semantic world.

Let us accept that the question “What is the time?” is relatively untransferable because its referent—the time—is highly specific to the “now” of its I-here-now and because it’s assumed code—mechanical clock-time—is not as universally available as the sun and the moon.

Now, accepting these restrictions, is it possible to imagine modes of textual presentation in which the same question could become eminently transferable? Obviously, the utterance could be part of a mathematical problem, explicitly conditioned by a series of co-textual astronomical observations and interpretative codes such that the reply should always be the same; or, as we have noted, the explicit co-text of Les Gommes uses similar principles to structure clock-time in
terms of an Oedipal code, able consistently to provide the question with far more than its isolated meaningfulness.

In such cases, transferability is significantly increased by textual presentation not only of a general semantic world, but also of the asymmetric “we” for whom the utterance is initially meaningful. Untransferability can thus be understood as a result of the absence of such contextual presentation; it ensues from dependence on unexpressed or implicit context. More metaphorically, if a text cannot be taken away from its owners, it is sometimes possible to convert the owners into signs and to transfer them along with the text.

3.2 Equivalence

Equivalence plays very important role in the transferability. Without equivalence it will hard to transfer the meaning in the very right way.

Once the text of source language L1 is not transferred or translated properly it will create problem in reader’s mind of target language L2. For this translator must be very keen and careful in selecting the equivalents.

3.2.1 Equivalence could be all things to all Theorists

Although descriptions of the relation between the input and output of translational work often refer to notions of equivalence, the term would appear to be the great empty sign of such exercises. Equivalence has been extensively used to define translation, but few writers have been prepared to define equivalence itself. Indeed, it is quite possible that the term in question means all things to all theorists: since it is usually taken to be the result of successful translating, its content as a theoretical term is probably nothing more or less than the theory in which successful translating is defined.
Equivalence thus perhaps means achieving whatever the ideal translator should set out to achieve. Yet this is a mere tautology: equivalence is supposed to define translation, but translation would then appear to define equivalence. One senses that something more substantial needs to be said about equivalence itself.

Historical research is of little avail here. The brief survey offered by Wilss (1982, 134-135) simply presents guesses suggesting that the English term “equivalence” entered translation studies from mathematics, that it was originally associated with research into machine translation, and that it has or should have a properly technical sense.

But Snell-Hornby has used comparative historical analysis to argue against the possibility of any such technical sense, claiming to have located some 58 different types of equivalence referred to in German translation studies (1986, 15). Moreover, even if one could locate substantial common factors underlying all these variants, there is surely no guarantee that history or etymology alone will lead to the most fruitful future definition. A slightly more creative approach is required.

Despite all the problems with historical usages of the term, despite recently fashionable attempts to ignore it altogether, one should believe that equivalence in its most unqualified form-definitionally ideal equivalence does indeed define translation. But to reach this conclusion, to discover what is being said but not heard, it is necessary to discard several false or inadequate notions of equivalence.

We must disregard the way structuralist linguistics once used the term to suggest a symmetry of “equal values” between discrete systems; we must turn to the economics of exchange in order to distinguish equivalence from assumptions of natural use values or functions; we must see how equivalence can actually operate
within a dynamic translational series based on the primacy of exchange value; and finally, we must appreciate that equivalence is not a predetermined relation that translators passively seek, but instead works as a transitory fiction that translators produce in order to have receivers somehow believe that translations have not really been translated. In all, if equivalence is ideally to define translation, we must take steps to redefine ideal equivalence.

3.2.2 Equivalence is directional and subject less

The following are fairly representative equivalence-based definitions of translation:

“Interlingual translation can be defined as the replacement of elements of one language, the domain of translation, by equivalent elements of another language, the range [of translation].” (A.G. Oettinger 1960, 110)

“Translation may be defined as follows: the replacement of textual material in one language (SL) by equivalent material in another language (TL).” (Catford 1965, 20)

“Translating consists in reproducing in the receptor language the closest natural equivalent of the source-language message.” (Nida and Taber 1969, 12; cf. Nida 1959, 33)

“[Translation] leads from a source-language text to a target-language text which is as close an equivalent as possible and presupposes an understanding of the content and style of the original.” (Wilss 1982, 62)

Many further definitions could be added in this vein (cf. Koller 1979, 186 ff.). But the main variants in any longer listing would tend to concern more the nature of what is supposed to be equivalent (“elements”, “textual material”, “functions”, “communicative effect”, etc.) than the nature of equivalence itself,
which, within this decidedly twentieth-century tradition, is simply assumed to exist.

Indeed, in some circles, the assumption is so amorphously present that one hesitates to question its grounding. Even Quine’s definition of indeterminacy, despite all its efforts explicitly to question contemporary presuppositions, feigns to be upset about the same text leading to different translations “which stand to each other in no plausible sort of equivalence relation however loose” (1960, 27).

But who told Quine that wholly determined translation should depend on equivalence? Is it not strange that equivalence thus appears in the definition of both what we know about translation (determinacy) and what we suspect we do not know (indeterminacy)? But what then is equivalence itself, however loose?

It might of course be assumed that the term means exactly what it says: a relation of equal value. But such a reading would contradict the similarly widespread although perhaps less obvious features we find that the term equivalence is commonly associated with the end result of translating as a one-way process occurring in an apparently subjectless place. Equivalence is directional and subjectless.

One should believe that these distinctive features are highly useful for the definition of translation. Moreover, their implicit asymmetry presents significant problems for certain less definite ideals like equivalence as an affair of “equal values”. The first of these problems is the nature of value itself.

3.2.3 Equivalence is Asymmetrical

Although “value” is generally not a technical term in contemporary translation studies, it does make frequent and prolonged appearances in Saussure’s *Cours de linguistique générale*, widely held to be one of the foundational texts of modern
linguistics and often cited in arguments against translatability. Saussure describes linguistic elements as having values corresponding to their mutual oppositions:

“Modern French mouton can have the same signification as English sheep but not the same value, and this for several reasons, particularly because in speaking of a piece of meat ready to be served at the table, English uses mutton and not sheep. The difference in value between sheep and mouton is due to the fact that sheep has beside it a second term while the French does not.” (1916,115)

Saussurean value is thus positional and relative within a fixed tongue, since “in language there are only differences without positive terms” (120). It is important to stress Saussure’s distinction between, on the one hand, “value” as the entire semantic potential left to an element by the presence or absence of neighboring terms, and on the other, “signification” as the particular use made of that element in a given situation.

This distinction is clear in the example of the chess game, where the value of the knight is described as its capacity to carry out any number of moves within the limits of certain rules, its signification then being the import of each individual move. So far, so good.

Since Saussurean value refers to the relative positions of elements within an entire tongue, the fact that different tongues divide semantic space in different ways theoretically denies the very possibility of different elements being of equal value.

Vendryes even considered equivalence to be contrary to the nature of the tongue as system, arguing that as soon as two elements become “equivalent” within the same system, one of them is forced to disappear (1923, 381).
It is then not surprising that Saussure’s synchronic linguistics excludes not only questions of equivalence but also all reference to one-way processes and to places of lesser dimensions than tongues. Saussure does not talk about translation. For example, he chooses not to tell us that the difference in value between “sheep” and “mutton” is due to the historical situation in which Anglo-Saxon servants presented what they called “sceap” to their Norman masters, who called the same object “moton”. The positional values of the terms were changed were exchanged as soon as the meat approached the master’s table and intercultural communication was established.

It is only through asymmetric situations like this—which clearly involve translation and quite massive material transfer (of meat, of armies), as indeed does Saussure’s description of the example—that the linguist has access to the comparable terms enabling him paradoxically to demonstrate that equal values (and thus “translation” itself) are strictly impossible. But the pertinent translation had taken place centuries before!

Equivalence can be defined in terms of exchange value, expressed as a relationship between texts and determined in the specific locus of the translator as a silent trader. This is what was being said but not heard.

The theory of translation is concerned with a certain types of relations between languages and is consequently a branch comparative linguistics. From the point of view of translation theory the distinction between synchronic and diachronic comparison is irrelevant.

Translation equivalence may be set up and translation performed between any pair of languages or dialects- related or unrelated and with one kind of spatial, temporal, social or other relationship between them.
Relation between languages can generally be regarded as two directional, though not always symmetrical. Translation, as a process, is always unidirectional, the replacement of textual material in one language (SL) by the equivalent textual material in another language (TL). It is always performed in a given direction from a source language (SL) L1 into a target language (TL) L2.

The equivalence factor occupies a major position in the translation theory. Catford (1965) views translation task as that of establishing target language (TL) equivalence for source language (SL) textual element. He sets out to a sign specific distinctive feature for each speech situation and evaluation, the translation equivalence in term of such feature as the ‘greater the number feature common to the better the translation’.

According to Halverson (1997), analogies between the equivalence concept and a concept of scientific knowledge as it is and has been studied with in the philosophy of science are highly informative in painting out the philosophical issues involved in equivalence, translation, and knowledge. He also believes that rather than dismissing the concept as ill-defined or imprecise, it is in the interest of the field of translation studies to consider the origins and manifestations of this ‘imprecision’ in order that we may be better informed and less inclined towards theoretical antagonism.

Therefore, many studies have been focused on the nature, interlingual and intertextual, empirical and theoretical notion of equivalence in recent years (Catford 1965, 1994, Pym 1992, Koller 1979, Toury 1980, Hutchins and Somers 1992, Arnold 1994). The domain of equivalents covers linguistic units such as morphemes, words, phrases, clauses, idioms and proverbs (Baker 1992). Through using finding equivalence strategies, the translators also attempt to improve the
chance of persuading their readers by making better their qualities of translation (Neubert 1985).

Amongst the grammatical devices which might cause problems in translation Baker focuses on number, tense and aspects, voice, person and gender. equivalence at word level, Baker (1992, p.26-42) proposes the following classification of strategies to solve non-equivalence at word level.

From the above it can be drawn that in the study of equivalence in translation shows how translators accurately render text in translation from source language (SL) into target language or vice versa. Therefore the translators, by finding equivalence in translation can show the tentative nature of their assertions, invite the readers, as intelligent individuals, to join and choose which translation is accurately render the thoughts, concepts and words of original text.

3.3 Dictionary based Translation
The best of example of dictionary based translation is machine translation. Machine translation uses a method based on dictionary entries, which means that the words will be translated as a dictionary does word by word, usually without much correlation of meaning between them.

Dictionary lookups may be done with or without morphological analysis or lemmatisation. While this approach to machine translation is probably the least sophisticated, dictionary-based machine translation is ideally suitable for the translation of long lists of phrases on the subsentential (i.e., not a full sentence) level, e.g. inventories or simple catalogs of products and services.

It can also be used to expedite manual translation, if the person carrying it out is fluent in both languages and therefore capable of correcting syntax and grammar.
When practicing a cross-language information retrieval, either the query or the target documents need to be translated. For obvious reasons, it usually is the query which gets to be put through a translator. There are many ways to try to find corresponding translations for the search keys but dictionary based translation is one of the most commonly used approaches. Dictionary based approaches are fairly easy to implement and they offer quite effective tools for retrieving translations for unknown search keys.

In its simplicity, dictionary based translation is basically translation with a help of a dictionary. When user inputs a query to a information retrieval system, the query is attempted to translate using a dictionary in recovering the desired translation. It is not surprising that there are many problems which one might encounter during this kind of process.

In dictionary based translation the words and phrases are translated on the basis of dictionary entries. But the biggest problem arises when a word / word string of a source language L1 have more than one meaning. When this word of L1 translated into target language L2 it forms ambiguity or loss of its real meaning.

**Problems in Dictionary based Translation (Machine Translation MT)**

3.3.1 Word missing in Dictionary entry

It is impossible to have a dictionary as extensive as there are words in any natural language. In addition to the practical limitations there are also theoretical ones. New words and word combinations are generated all the time in all fields of the human conversation.

Even if all of them never end up in to the common use, the task is still, quite impossible. Words missing from a dictionary, or words out-of-vocabulary, as the problem is also referred, is a huge problem for dictionary based translation
systems. If there is no word in the dictionary for an unknown search key, the key is usually left untranslated.

The most important categories of untranslatable search keys are new compound words, proper names and their spelling variants and special terms.

### 3.3.2 Inflected words

In many languages the words are marked to reflect their grammatical information such as gender, number or tense, this is called inflection. Inflection can occur for example in verbs, nouns and adjectives. What makes this a problem in dictionary based translation, is that inflected words are not listed in dictionaries in their inflected form but in their base form. Detecting inflected words requires its own tools which include for example morphological analysis.

### 3.3.3 Phrases

Phrases like *the black dog* or *full of tears* might be found from a dictionary but longer and more complicated phrases like *the white house at the end of the street* are not listed anywhere. Phrases are not as problematic as usual for languages like German or Finnish where multi-word expressions are rather compound words than phrases. But in a case of English, like above, or for example in Urdu it will get more complicated. From the point of view compound words are more desirable because their decomposition is easier than phrase identification.

### 3.3.4 Lexical Ambiguity

One major problem is the natural lexical ambiguity of source and target languages. In every language, there are words which have multiple senses. As an example, the English word *string* has 17 different meanings according to Word Net. To make everything even more difficult, these 17 meanings can be (in this case) either nouns or verbs. Even if the word *string* was found in a dictionary it would have 17
senses, how could a machine know which one is the correct translation? This kind of word sense identification requires heavy examination of the word context.

For instance when an English language L1 word string ‘movement in engine’ translated into Urdu/Hindi language L2 as ‘engine me aandolan’ means ‘revolution in engine’ instead of ‘speed in engine’ the whole translation work lost its meaning.

So from the above it can be said that in dictionary based translation the translated work from source language L1 to target language L2 has become much more complicated and it also opens the doors for ambiguity.

3.4 Selection of ‘Keywords’

In linguistics and especially in translation, selection of keywords which carries the functional load (also referred to as communicative load) refers to the importance of the certain features in making distinctions in a language. In other words, how hard would it be to guess and identity of a word in context which has communicative load?

A term or a phrase used by a searcher to find information on a particular topic is called a "keyword." "Keyword selection" is the process of scrutinizing different prospective keywords to select the right ones for transferability. It is the stepping stone to success in the translation world. Find the right keyword, and one would strike gold. Make a mistake, and translator just punched himself in the face.

3.4.1 Keyword Analysis or Keyword Research

We move on with the first step in keyword selection. This is keyword analysis. To start with, we need at least one keyword. So we determine a primary keyword. This should be very specific. In translation studies when source text L1 translated into target text L2 these keywords play an essential role. By selecting these
keywords or words which have communicative load translated, the translation work L2 quite near to the source text L2.

Translation theory shares a number of concerns with what is commonly called communication theory. Perhaps the most important observation which the communication theorists have produced for translators is the recognition that every act of communication has three dimensions: Speaker (or author), Message, and Audience. The more we can know about the original author, the actual message produced by that author, and the original audience, the better acquainted we will be with that particular act of communication. An awareness of this tripartite character of communication can be very useful for interpreters.

In communication the degree of ability which a receptor has to understand a message, channel capacity is conditioned both by the receptor's personnel qualities and by his cultural background and is a function of the amount of information which the receptor has in common with the author. The narrower the channel capacity, the more redundancy needs to be introduced to lighten the communication load.

3.5 Disambiguation

Word - sense removal of ambiguity or disambiguation concerns finding a suitable translation when a word can have more than one meaning. The problem was first raised in the 1950s by Yehoshua Bar-Hillel. He pointed out that without a "universal encyclopedia", a machine would never be able to distinguish between the two meanings of a word. Today there are numerous approaches designed to overcome this problem. They can be approximately divided into "shallow" approaches and "deep" approaches.
Shallow approaches assume no knowledge of the text. They simply apply statistical methods to the words surrounding the ambiguous word. Deep approaches presume a comprehensive knowledge of the word. So far, shallow approaches have been more successful.

The late Claude Piron, a long-time translator for the United Nations and the World Health Organization, wrote that machine translation, at its best, automates the easier part of a translator's job; the harder and more time-consuming part usually involves doing extensive research to resolve ambiguities in the source text, which the grammatical and lexical exigencies of the target language require to be resolved:

Why does a translator need a whole workday to translate five pages, and not an hour or two? About 90% of an average text corresponds to these simple conditions. But unfortunately, there's the other 10%. It's that part that requires six or more hours of work. There are ambiguities one has to resolve. For instance, the author of the source text, an Australian physician, cited the example of an epidemic which was declared during World War II in a "Japanese prisoner of war camp". Was he talking about an American camp with Japanese prisoners or a Japanese camp with American prisoners? The English has two senses. It's necessary therefore to do research, maybe to the extent of a phone call to Australia.

The ideal deep approach would require the translation software to do all the research necessary for this kind of disambiguation on its own; but this would require a higher degree of Artificial Intelligence than has yet been attained.

A shallow approach which simply guessed at the sense of the ambiguous English phrase that Piron mentions (based, perhaps, on which kind of prisoner-of-war
camp is more often mentioned in a given corpus) would have a reasonable chance of guessing wrong fairly often. A shallow approach that involves "ask the user about each ambiguity" would, by Piron's estimate, only automate about 25% of a professional translator's job, leaving the harder 75% still to be done by a human.

So in the light of the above to remove the ambiguity from source text L1 to target text L2 manual human interference is necessary. Knowledge of the cultures is essential. Knowledge of the context is important to come over the ambiguity or to make text disambiguous.

3.6 Word based Disambiguation

Word based disambiguation or Word sense disambiguation is the process of identifying the sense of a word in a sentence.

In translation studies, word sense disambiguation (WSD) is an open problem of translating the source language L1 to target language L2, which governs the process of identifying which sense of a word that is the meaning used in a sentence, when the word has multiple meanings which creates the situation known as polysemy. Especially this situation is worst when translation job is done by machine i.e. Machine Translation (MT)

The solution to this problem will impact other related writing, such as discourse, improving relevance of search engines, anaphora resolution, coherence, inference and others.

Research has progressed steadily to the point where WSD systems achieve sufficiently high levels of accuracy on a variety of word types and ambiguities. A rich variety of techniques have been researched, from dictionary-based methods that use the knowledge encoded in lexical resources, to supervised machine learning methods in which a classifier is trained for each distinct word on a corpus
of manually sense-annotated examples, to completely unsupervised methods that cluster occurrences of words, thereby inducing word senses. Among these, supervised learning approaches have been the most successful algorithms to date.

A disambiguation process requires two strict things: a dictionary to specify the senses which are to be disambiguated and a corpus of language data to be disambiguated (in some methods, a training corpus of language examples is also required). WSD task has two variants: "lexical sample" and "all words" task. The former comprises disambiguating the occurrences of a small sample of target words which were previously selected, while in the latter all the words in a piece of running text need to be disambiguated. The latter is deemed a more realistic form of evaluation, but the corpus is more expensive to produce because human annotators have to read the definitions for each word in the sequence every time they need to make a tagging judgment, rather than once for a block of instances for the same target word.

To give a hint how all this works, consider an example of the distinct senses that exist for the (written) word "peer":

peer ne peer ko dawat di ‘peer (sufi) invited on monday’
peer = sufi
peer = 1st day of the week

To a human, it is obvious that in the sentence using the word "peer", as in the former sense above and in the second sentence, the word "peer" is being used as in the latter sense below. Developing algorithms to replicate this human ability can often be a difficult task.
3.6.1 Difficulties in Disambiguation

**Differences between dictionaries**

One problem with word sense disambiguation is deciding what the senses are. In cases like the word *bass* above, at least some senses are obviously different. In other cases, however, the different senses can be closely related (one meaning being a metaphorical or metonymic extension of another), and in such cases division of words into senses becomes much more difficult. Different dictionaries and thesauruses will provide different divisions of words into senses. One solution some researchers have used is to choose a particular dictionary, and just use its set of senses. Generally, however, research results using broad distinctions in senses have been much better than those using narrow. However, given the lack of a full-fledged coarse-grained sense inventory, most researchers continue to work on fine-grained WSD.

Most research in the field of WSD is performed by using Word Net as a reference sense inventory for English. Word Net is a computational lexicon that encodes concepts as synonym sets (e.g. the concept of car is encoded as \{ car, auto, automobile, machine, motorcar \}). Other resources used for disambiguation purposes include Roget's Thesaurus and Wikipedia.

**Part-of-speech tagging**

In any real test, part-of-speech tagging and sense tagging are very closely related (it concerns only some languages, e.g. English, Urdu) with each potentially making constraints to each other. And the question whether these tasks should be kept together or decoupled is still not unanimously resolved, but recently scientists incline to test these things separately (e.g. in the Senseval competitions parts of speech are provided as input for the text to disambiguate).
It is instructive to compare the word sense disambiguation problem with the problem of part-of-speech tagging. Both involve disambiguating or tagging with words, be it with senses or parts of speech. However, algorithms used for one do not tend to work well for the other, mainly because the part of speech of a word is primarily determined by the immediately adjacent one to three words, whereas the sense of a word may be determined by words further away. The success rate for part-of-speech tagging algorithms is at present much higher than that for WSD, state-of-the-art being around 95% accuracy or better, as compared to less than 75% accuracy in word sense disambiguation with supervised learning. These figures are typical for English & Urdu and may be very different from those for other languages.

**Inter-judge variance**

Another problem is inter-judge variance. WSD systems are normally tested by having their results on a task compared against those of a human. However, while it is relatively easy to assign parts of speech to text, training people to tag senses is far more difficult. While users can memorize all of the possible parts of speech a word can take, it is impossible for individuals to memorize all of the senses a word can take. Moreover, humans do not agree on the task at hand give a list of senses and sentences, and humans will not always agree on which word belongs in which sense.

Thus, a computer cannot be expected to give better performance on such a task than a human (indeed, since the human serves as the standard, the computer being better than the human is incoherent), so the human performance serves as an upper bound. Human performance, however, is much better on coarse-grained than fine-grained distinctions, so this again is why research on coarse-grained distinctions has been put to test in recent WSD evaluation exercises.
Common sense

Some Artificial Intelligence researchers like Douglas Lenat argue that one cannot parse meaning from words without some form of common sense ontology. For example, comparing two these sentences:

"Fatma and Jameela are sisters." — (they are sisters of each other).
"Fatma and Jameela are mothers." — (each is independently a mother).

To properly identify senses of words one must know common sense facts. Moreover, sometimes the common sense is needed to disambiguate such words like pronouns in case of having anaphoras or cataphoras in the text.

Sense inventory and algorithms' task-dependency

A task-independent sense inventory is not a coherent concept: each task requires its own division of word meaning into senses relevant to the task. For example, the ambiguity of 'mouse' (animal or device) is not relevant in English-French machine translation, but is relevant in information retrieval. In case of Urdu it will be the problem. The opposite is true of 'river', which requires a choice in French (fleuve 'flows into the sea', or rivière 'flows into a river').

Also, completely different algorithms might be required by different applications. In machine translation, the problem takes the form of target word selection. Here the "senses" are words in the target language, which often correspond to significant meaning distinctions in the source language (aaya could translate to Urdu aaya ‘governess' or aaya 'come').

Discreteness of senses

Finally, the very notion of "word sense" is slippery and controversial. Most people can agree in distinctions at the coarse-grained homograph level (e.g., pen as writing instrument or enclosure), but go down one level to fine-grained polysemy,
and disagreements arise. For example, in Senseval-2, which used fine-grained sense distinctions, human annotators agreed in only 85% of word occurrences. Word meaning is in principle infinitely variable and context sensitive. It does not divide up easily into distinct or discrete sub-meanings.

Lexicographers frequently discover in corpora loose and overlapping word meanings, and standard or conventional meanings extended, modulated, and exploited in a bewildering variety of ways. The art of lexicography is to generalize from the corpus to definitions that evoke and explain the full range of meaning of a word, making it seem like words are well-behaved semantically.

However, it is not at all clear if these same meaning distinctions are applicable in computational applications, as the decisions of lexicographers are usually driven by other considerations. Recently, a task - named lexical substitution - has been proposed as a possible solution to the sense discreteness problem. The task consists of providing a substitute for a word in context that preserves the meaning of the original word (potentially, substitutes can be chosen from the full lexicon of the target language, thus overcoming discreteness).

3.6.2 Approaches and Methods
As in all natural language processing, there are two main approaches to WSD are deep approaches and shallow approaches.

Deep approaches presume access to a comprehensive body of world knowledge. Knowledge, such as "you can go fishing for a type of fish, but not for low frequency sounds" and "songs have low frequency sounds as parts, but not types of fish", is then used to determine in which sense the word is used.

These approaches are not very successful in practice, mainly because such a body of knowledge does not exist in a computer-readable format, outside of very limited
domains. However, if such knowledge did exist, then deep approaches would be much more accurate than the shallow approaches. Also, there is a long tradition in computational linguistics, of trying such approaches in terms of coded knowledge and in some cases; it is hard to say clearly whether the knowledge involved is linguistics or world knowledge.

The first attempt was that by Margaret Masterman and her colleagues, at the Cambridge Language Research Unit in England, in the 1950s. This attempt used as data a punched-card version of Roget's Thesaurus and its numbered "heads", as an indicator of topics and looked for repetitions in text, using a set intersection algorithm. It was not very successful, but had strong relationships to later work, especially Yarowsky’s machine learning optimisation of a thesaurus method in the 1990s.

Shallow approaches don't try to understand the text. They just consider the surrounding words, using information such as "if peer has words days or weeks nearby, it probably is in the day sense; if peer has the words person or human nearby, it is probably in the name sense." These rules can be automatically derived by the computer, using a training corpus of words tagged with their word senses. This approach, while theoretically not as powerful as deep approaches, gives superior results in practice, due to the computer's limited world knowledge. However, it can be confused by sentences like dawa pili hai ‘I have taken the medicine and medicine is yellow’ which contains the word bark near both color and take.

There are four conventional approaches to word based disambiguation or word sense disambiguation (WSD):
- **Dictionary- and knowledge-based methods**: These rely primarily on dictionaries, thesauri, and lexical knowledge bases, without using any corpus evidence.

- **Supervised methods**: These make use of sense-annotated corpora to train from.

- **Semi-supervised or minimally-supervised methods**: These make use of a secondary source of knowledge such as a small annotated corpus as seed data in a bootstrapping process, or a word-aligned bilingual corpus.

- **Unsupervised methods**: These eschew (almost) completely external information and work directly from raw un-annotated corpora. These methods are also known under the name of word sense discrimination.

Almost all these approaches normally work by defining a window of N content words around each word to be disambiguated in the corpus, and statistically analyzing those N surrounding words. Two shallow approaches used to train and then disambiguate are Naïve Bayes classifiers and decision trees. In recent research, kernel-based methods such as support vector machines have shown superior performance in supervised learning. Graph-based approaches, that currently achieve performance close to the state of the art, have also gained much attention from the research community.

### 3.6.2 (i) Dictionary- and knowledge-based methods

The Lesk algorithm is the seminal dictionary-based method. It is based on the hypothesis that words used together in text are related to each other and that the relation can be observed in the definitions of the words and their senses. Two (or more) words are disambiguated by finding the pair of dictionary senses with the greatest word overlap in their dictionary definitions. For example, when
disambiguating the words in "paye", the definitions of the appropriate senses both include the words *got* and *legs* (at least in one dictionary).

An alternative to the use of the definitions is to consider general word-sense relatedness and to compute the semantic similarity of each pair of word senses based on a given lexical knowledge base such as Word Net.

Graph-based methods reminiscent of spreading activation research of the early days of Artificial Intelligence research have been applied with some success. More complex graph-based approaches have been shown to perform almost as well as supervised methods or even outperforming them on specific domains. Recently, it has been reported that simple graph connectivity measures, such as degree, perform state-of-the-art WSD in the presence of a sufficiently rich lexical knowledge base. Also, automatically transferring knowledge in the form of semantic relations from Wikipedia to WordNet has been shown to boost simple knowledge-based methods, enabling them to rival the best supervised systems and even outperform them in a domain-specific setting.

The use of selectional preferences (or selectional restrictions) is also useful, for example, knowing that one typically cooks food, one can disambiguate the word bass in "I am cooking basses" (i.e., it's not a musical instrument).

**3.6.2 (ii) Supervised methods**

Supervised methods are based on the assumption that the context can provide enough evidence on its own to disambiguate words (hence, world knowledge and reasoning are deemed un-necessary). Probably every machine learning algorithm going has been applied to WSD, including associated techniques such as feature selection, parameter optimization, and ensemble learning. Support Vector Machines and memory-based learning have been shown to be the most successful approaches, to date, probably because they can cope with the high-dimensionality
of the feature space. However, these supervised methods are subject to a new knowledge acquisition bottleneck since they rely on substantial amounts of manually sense-tagged corpora for training, which are laborious and expensive to create.

3.6.2 (iii) Semi-supervised methods

Because of the lack of training data, many word sense disambiguation algorithms use semi-supervised learning, which allows both labeled and unlabeled data. The Yarowsky algorithm was an early example of such an algorithm. It uses the ‘One sense per collocation’ and the ‘One sense per discourse’ properties of human languages for word sense disambiguation. From observation, words tend to exhibit only one sense in most given discourse and in a given collocation.

The bootstrapping approach starts from a small amount of seed data for each word: either manually-tagged training examples or a small number of surefire decision rules (e.g., 'din = day' in the context of 'peer = Monday' almost always indicates 1st day of the week).

The seeds are used to train an initial classifier, using any supervised method. This classifier is then used on the untagged portion of the corpus to extract a larger training set, in which only the most confident classifications are included. The process repeats, each new classifier being trained on a successively larger training corpus, until the whole corpus is consumed, or until a given maximum number of iterations are reached.

Other semi-supervised techniques use large quantities of untagged corpora to provide co-occurrence information that supplements the tagged corpora. These techniques have the potential to help in the adaptation of supervised models to different domains.
Also, an ambiguous word in one language is often translated into different words in a second language depending on the sense of the word. Word-aligned bilingual corpora have been used to infer cross-lingual sense distinctions, a kind of semi-supervised system.

3.6.2 (iv) Unsupervised methods

Unsupervised learning is the greatest challenge for WSD researchers. The underlying assumption is that similar senses occur in similar contexts, and thus senses can be induced from text by clustering word occurrences using some measure of similarity of context, a task referred to as word sense induction or discrimination. Then, new occurrences of the word can be classified into the closest induced clusters/senses. Performance has been lower than other methods, above, but comparisons are difficult since senses induced must be mapped to a known dictionary of word senses. If a mapping to a set of dictionary senses is not desired. Cluster-based evaluations (including measures of entropy and purity) can be performed.

Alternatively, word sense induction methods can be tested and compared within an application. For instance, it has been shown that word sense induction improves Web search result clustering by increasing the quality of result clusters and the degree diversification of result lists. It is hoped that unsupervised learning will overcome the knowledge acquisition bottleneck because they are not dependent on manual effort.

3.6.2 (v) Other approaches

Few more other approaches may vary differently in their methods to work on word based disambiguation:

- Identification of dominant word senses;
- Domain-driven disambiguation;
WSD using Cross-Lingual Evidence;

The knowledge acquisition bottleneck is perhaps the major impediment to solve the Word based Disambiguation problem. Unsupervised methods rely on knowledge about word senses, which is barely formulated in dictionaries and lexical databases. Supervised methods depend crucially on the existence of manually annotated examples for every word sense, a requisite that can so far be met only for a handful of words for testing purposes.

Obviously knowledge is a fundamental component of Word based Disambiguation. Knowledge sources provide data which are essential to associate senses with words. They can vary from corpora of texts, either unlabeled or annotated with word senses, to machine-readable dictionaries, thesauri, glossaries, ontologies, etc. They can be classified as follows:

1. Structured:
   - Thesauri
   - Machine-readable dictionaries (MRDs)
   - Ontologies

2. Unstructured:
   - Corpora: raw corpora and sense-annotated corpora
   - Collocation resources
   - Other resources (such as word frequency lists, stop lists, domain labels, etc.)

3.7 The Mutual Information Leading to Disambiguation

Mutual information is one of many quantities that measures how much one random variable tells us about another. It is a dimensionless quantity with (generally) units of bits, and can be thought of as the reduction in uncertainty about one random variable given knowledge of another. High mutual information
indicates a large reduction in uncertainty; low mutual information indicates a small reduction; and zero mutual information between two random variables means the variables are independent.

3.7.1 Definition

For two discrete variable X and Y whose joint probability distribution is \( P_{xy}(x,y) \), the mutual information between them, denoted \( I(X; Y) \), is given by (Shannon and Weaver, 1949; Cover and Thomas, 1991)

\[
I(X; Y) = \sum_{x,y} P_{xy}(x,y) \log \frac{P_{xy}(x,y)}{P_x(x)P_y(y)} = E_{p_{xy}} \log \frac{P_{xy}}{P_x P_y}. \tag{1}
\]

Fig 3.1: Mutual Information between X & Y
Here \( P_X(x) \) and \( P_Y(y) \) are the marginals: 
\[
P_X(x) = \sum_y P_{XY}(x, y)
\]
and 
\[
P_Y(y) = \sum_x P_{XY}(x, y)
\]
and \( E_P \) is the expected value over the distribution \( P \).

The focus here is on discrete variables, but most results derived for discrete variables extend very naturally to continuous ones – one simply replaces sums by integrals. One should be aware, though, that the formal replacement of sums by integrals hides a great deal of subtlety, and, for distributions that are not sufficiently smooth, may not even work.

### 3.7.2 Interpretation

To understand what \( I(X; Y) \) actually means, we first need to define entropy and *conditional entropy*.

Qualitatively, entropy is a measure of uncertainty – the higher the entropy, the more uncertain one is about a random variable. This statement was made quantitative by Shannon. He postulated that a measure of uncertainty of a random variable \( X \) should be a continuous function of its probability distribution \( P_X(x) \) and should satisfy the following conditions:

- It should be maximal when \( P_X(x) \) is uniform, and in this case it should increase with the number of possible values \( X \) can take;
- It should remain the same if we reorder the probabilities assigned to different values of \( X \);
- The uncertainty about two independent random variables should be the sum of the uncertainties about each of them.

He then showed that the only measure of uncertainty that satisfies all these conditions is the entropy, defined as
Although not particularly obvious from this equation, $H(X)$ has a very concrete interpretation: Suppose $x$ is chosen randomly from the distribution $P_X(i)$, and someone who knows the distribution $P_X(i)$ is asked to guess which $x$ was chosen by asking only yes/no questions. If the guesser uses the optimal question-asking strategy, which is to divide the probability in half on each guess by asking questions like "is $x$ greater than $x_0$?", then the average number of yes/no questions it takes to guess $x$ lies between $H(X)$ and $H(X)+1$ (Cover and Thomas, 1991). This gives quantitative meaning to "uncertainty": it is the number of yes/no questions it takes to guess a random variable, given knowledge of the underlying distribution and taking the optimal question-asking strategy.

The conditional entropy is the average uncertainty about $X$ after observing a second random variable $Y$, and is given by

$$H(X|Y) = \sum_y P_Y(y) \left[ -\sum_x P_{X|Y}(x|y) \log \left( P_{X|Y}(x|y) \right) \right] = E_{P_Y} \left[ -E_{P_{X|Y}} \log P_{X|Y} \right] \quad \text{(3)}$$

where $P_{X|Y}(x|y)(= P_{X|Y}(x,y)/P_Y(y))$ is the conditional probability of $x$ given $y$.

With the definitions of $H(X)$ and $H(X|Y)$, Eq. (1) can be written

$$I(X,Y) = H(X) - H(X|Y). \quad \text{(4)}$$

Mutual information is therefore the reduction in uncertainty about variable $X$, or the expected reduction in the number of yes/no questions needed to guess $X$ after
observing $Y$. Note that the yes/no question interpretation even applies to continuous variables: although it takes an infinite number of questions to guess a continuous variable, the difference in the number of yes/no questions it takes to guess $X$ before versus after observing $Y$ may be finite, and is the mutual information. While problems can arise when going from discrete to continuous variables (subtracting infinities is always dangerous), they rarely do in practice; see (Gray, 1990).

### 3.7.3 Properties

- Mutual information is intimately related to the Kullback-Leibler divergence (Cover and Thomas, 1991), a very natural measure of the ‘distance’ between two distributions. It is defined as follows: for any two distributions $P(z)$ and $Q(z)$,

$$D_{KL}(P(z)||Q(z)) \equiv \sum_z P(z) \log \left( \frac{P(z)}{Q(z)} \right).$$

Distance is in quotes because the Kullback-Leibler divergence is not a true distance: it is not symmetric, and it does not obey the triangle inequality (Cover and Thomas, 1991). It is not hard to show that $D_{KL}(P(z)||Q(z))$ is non-negative, and zero if and only if $P(z) = Q(z)$.

From Eq. (1), it is easy to see that the mutual information is just the Kullback-Leibler distance between the joint distribution, $P_{XY}(x,y)$, and the product of the independent ones, $P_X(x)P_Y(y)$,

$$I(X;Y) = D_{KL}(P_{XY}(x,y)||P_X(x)P_Y(y)).$$

Thus, another way to think about mutual information is that it is a measure of how close the true joint distribution of $X$ and $Y$ is to the independent joint
distribution. Because the Kullback-Leibler distance, and thus the mutual information, is zero if and only if $P_{XY}(x, y) = P_X(x)P_Y(y)$, it follows that the mutual information captures all dependencies between random variables, not just, say, second order ones, as captured by the covariance.

- Mutual information is symmetric: $I(X; Y) = I(Y; X)$. This follows directly from the definition, Eq. (1).

- Mutual information is additive for independent variables. More quantitatively, if $P_{XWYZ}(x,y,w,z) = P_{XY}(x,y)P_{WZ}(w,z)$, then
  \[ I(X,W;Y,Z) = I(X;Y) + I(W;Z). \]
  This follows easily from the definition of the mutual information. It's also a property that information should have. For example, suppose $X$ and $Y$ are the cost of a bottle of juice and how it tastes, and $W$ and $Z$ are the height and weight of tree squirrels. If cost provides 1 bit of information about taste and height provides 2 bits of information about weight, then it makes sense that cost + weight should provide 3 bits of information about taste + height. Assuming, of course, that squirrels (and their weights) do not influence juices (and their prices).

- The Data Processing Inequality (DPI) states, loosely, that post-processing cannot increase information. More quantitatively, consider two random variables, $X$ and $Y$, whose mutual information is $I(X,Y)$. Now consider a third random variable, $Z$, that is a (probabilistic)function of $Y$ only. The only qualifier means $P_{Z|XY}(z|x,y) = P_{Z|Y}(z|y)$, which in turn implies that $P_{X|YZ}(x|y,z) = P_{X|Y}(x|y)$, as is easy to show using Bayes' theorem. The DPI states that $Z$ cannot have more information about $X$ than $Y$ has about $X$; that is,
This inequality, which again is a property information *should* have, is easy to prove,

\[ I(X;Z) = H(X) - H(X|Z) \leq H(X) - H(X|Y,Z) = H(X) - H(X|Y) = I(X;Y). \]

The first equality follows from Eq. (4); the inequality follows because conditioning on an extra variable (in this case \(Y\) as well as \(Z\)) can only decrease entropy, and the second to last equality follows because \(P_{X|YZ}(x|y,z) = P_{X|Y}(x|y)\).

### 3.7.4 The channel coding theorem

Mutual Information is just one way among many of measuring how related two random variables are. However, it is a measure ideally suited for analyzing communication channels.

Abstractly, a communication channel can be visualized as a transmission medium which receives an input \(x\) and produces an output \(y\). If the channel is *noiseless*, the output will be equal to the input. However, in general, the transmission medium is noisy and an input \(x\) is converted to an output \(y\) with probability \(P_{Y|X}(y|x)\).

Given a communication channel, one can transmit any message \(s\) from a set of \(M\) possible messages by performing the following three steps:

1. To each message \(s\) assign a string \(x(s) = (x_1, \ldots, x_n)\) of length \(n\). Each \(x(s)\) is called a *codeword*. The deterministic mapping from the set of messages to the set of codewords is called the *encoding function*. 


2. To transmit \( s \), transmit the corresponding string \( x(s) \) over the channel. This yields an output string \( y \), also of length \( n \). For the so-called memoryless channels, each \( x_i \) in the input string is mapped to an output \( y_i \) with probability \( P_{Y|X}(y|x) \), independent of the other \( x_j \).

3. The output string \( y \) is used to recover the transmitted message, using a deterministic decoding function. The decoding function maps each \( y \) to one symbol \( s' \).

There are two elements of a communication channel that should be emphasized. First, the number of messages is typically much less than the number of possible messages. For example, if the \( x_i \) were binary, for which the number of possible messages is \( 2^n \), then one would typically have \( M \ll 2^n \) (see 1). Second, for each message, the \( x_i \) are chosen randomly, and independently, from a distribution denoted, as usual, \( P_X(x) \). When designing a channel, then, one has control over only two quantities: \( M \) and \( P_X(x) \). In fact, one usually adjusts \( P_X(x) \) to make the number of messages, \( M \), large while at the same time keeping the error rate - the rate at which messages are decoded incorrectly - small. Note that the conditional distribution, \( P_{Y|X}(y|x) \), is a physical property of the channel itself, and thus not under the control of the designer.

The three steps of the communication channel are illustrated as

\[
\begin{align*}
&\text{s} \xrightarrow{\text{Encoding}} x_1 x_2 \ldots x_n \rightarrow \left[ \text{noisy channel: } P_{Y|X}(y|x) \right] \rightarrow y_1 y_2 \ldots y_n \xrightarrow{\text{Decoding}} s'.
\end{align*}
\]

One would like \( s' \) to be the same as \( s \), but, because of noise, there is no way to absolutely guarantee this. However, the channel coding theorem (Shannon and Weaver, 1949) states that one can come close, so long as \( M \) is not too large. Moreover - and this is where the link between mutual information and
communication channels arises - the maximum number of messages that can be transmitted almost error-free is a function of the mutual information between $X$ and $Y$.

### 3.7.5 Variations of mutual information

Several variations on mutual information have been proposed to suit various needs. Among these are normalized variants and generalizations to more than two variables.

#### Metric

Many applications require a metric, that is, a distance measure between points. The quantity

$$d(X,Y) = H(X,Y) - I(X;Y) = H(X) + H(Y) - 2I(X;Y) = H(X | Y) + H(Y | X)$$

satisfies the basic properties of a metric; most importantly, the triangle inequality, but also non-negativity, indiscernability and symmetry. This distance metric is also known as the Variation of information.

Since one has $d(X,Y) \leq H(X,Y)$, a natural normalized variant is

$$D(X,Y) = d(X,Y)/H(X,Y) \leq 1.$$ 

The metric $D$ is a universal metric, in that if any other distance measure places $X$ and $Y$ close-by, then the $D$ will also judge them close.

A set-theoretic interpretation of information (see the figure for Conditional entropy) shows that

$$D(X,Y) = 1 - I(X;Y)/H(X,Y) = 1 - H(X \cap Y)/H(X \cup Y)$$
which is effectively the Jaccard distance between X and Y.

**Conditional mutual information**

Sometimes it is useful to express the mutual information of two random variables conditioned on a third, which can be simplified as

\[ I(X; Y | Z) = \sum_{z \in Z} \sum_{y \in Y} \sum_{x \in X} p_{X,Y,Z}(x, y, z) \log \frac{p_Z(z)p_{X,Y,Z}(x, y, z)}{p_{X,Z}(x, z)p_{Y,Z}(y, z)}. \]

Conditioning on a third random variable may either increase or decrease the mutual information, but it is always true that

\[ I(X; Y | Z) \geq 0 \]

for discrete, jointly distributed random variables X, Y, Z. This result has been used as a basic building block for proving other inequalities in information theory.

**Multivariate mutual information**

Several generalizations of mutual information to more than two random variables have been proposed, such as total correlation and interaction information. If Shannon entropy is viewed as a signed measure in the context of information diagrams, as explained in the article Information theory and measure theory, then the only definition of multivariate mutual information that makes sense is as follows:

\[ I(X_1; X_n) = H(X_n) \]

and for \( n > 1 \),

\[ I(X_1; \ldots; X_n) = I(X_1; \ldots; X_{n-1}) - I(X_1; \ldots; X_{n-1}|X_n), \]

where (as above) we define

\[ I(X_1; \ldots; X_{n-1}|X_n) = \mathbb{E}_{X_n}(I(X_1; \ldots; X_{n-1})|X_n). \]

This definition of multivariate mutual information is identical.
3.8 Conclusion

To conclude from the above, it can be asked that, exactly what transferability is. For the purposes of translation studies, the privileged object of transferability is the text, independently of whatever meaning, information, message or signification might have been attributed to that text prior for transferability.

One of the amazing things in translation studies is equivalence. To find the out the equivalence for target text L2 is very difficult. Something could not be detached from cultures of one’s society or area or sect or religion or country. So it is very hard to get cultured words equivalence.

For example equivalence words of sindoor, abaya, etc in English language is hard to get because their, they don’t exist, until and unless culture is transferred they are not easy to be understood. These words are directly related to the culture of the particular society/ religion. They are cultured words because cultures are directly related to them. This happened mainly in machine translation where everything is translated on the basis of dictionary, because dictionary based translation leads to the lost of real sense of words lost and it ends with many errors.

The keyword selection is also a very important factor in understanding and translating the text from source language L1 to target language L2. If translator gets the keyword of the source language L1, it becomes easy to translate into target language and it leads into error free translation.

One more important thing in transferability is the removal of ambiguity. Word - sense removal of ambiguity or disambiguation concerns finding a suitable translation when a word can have more than one meaning. Today there are numerous approaches designed to overcome this problem. They can be approximately divided into "shallow" approaches and "deep” approaches.
Therefore to remove the ambiguity from source text L1 to target text L2 manual human interference is necessary. Knowledge of the cultures is essential and comprehension of the context is very important to make text disambiguous

Furthermore, Mutual information also plays an important role in removal of the ambiguity or to make text disambiguous. Mutual information leading to disambiguation is one of many quantities that measures how much one random piece of information tells us about another. It can be thought of as the reduction in uncertainty about one random variable given knowledge of another. High mutual information indicates a large reduction in uncertainty; low mutual information indicates a small reduction; and zero mutual information between two random variables means the variables are independent.

Therefore in good translation there are not only words which have to be transferred from source language L1 to target language L2, but the cultures and their attached information should also be transferred.