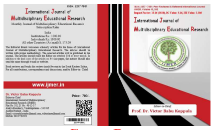
Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:6(1), June 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

VISUAL TRANSLATION: A LINGUAL SURVEY TO CUT THROUGH THE LANGUAGE WALLS

Dr. Bandar Sadanandam
Faculty, Department of English, P.G. Centre, Kollapur Palamuru University, Mahabubnagar. Telangana, India

Abstract:

Machine translation has revolutionized the world set on regional and language barriers that so often intrude the flow of communication. Homogeneous translation of texts over the stretch of languages has been the research focus of many computational linguists, but what about signs and symbols that have not only been ruling human existence from centuries but also have evolved the mammoth list of languages? However the interpretation of these lingual symbols often fails to communicate the real sense to the global society that has furnished on broken verses of language diversity. A far stretching voidance appears when we try to relate what we see, to what we interpret. So what if heterogeneous translation rises up its gears, leading to translation of visual information to understandable language texts? The paper ponders over the implementation of visual translation through computational generation of referring expressions, which would rather pave way to eliminate language constraints, thus polishing the communicational revolution that has set itself with the machines.

Keywords: Computation, Linguistics, Visual, Translation, Language

Introduction:

Much research in translation studies indicates that translated texts have unique characteristics that set them apart from original texts. Known as *translationese*, translated texts (in any language) constitute genre, or a dialect, of the target language, this reflects both artifacts of the translationprocess and traces of the original language from which the texts were translated. Among the better-known properties of translationese are *simplification* and *explicit*. Translated texts tend to be shorter, to have lower type/token ratio, and to use certain discourse markers more frequently than original texts. Interestingly, translated texts are so markedly different from original ones that automatic classification can identify them with very high accuracy. Contemporary statistical machine translation (SMT) systems use parallel corpora to train *translation models* that reflect source- and target-language phrase correspondences. Typically, SMT systems ignore the direction of translation of the parallel corpus. Given the unique properties of translationese, which operate asymmetrically from source to target language, it is reasonable to assume that this direction may affect the quality of the translation. Recently, Kurokawa, Goutte, and Isabelle (2009) showed that this is indeed the case. They trained a system to translate between French and English (and vice versa) using a French-translated-to-English parallel corpus, and then an English translated-to-French one. They find that in translating into French the latter parallel corpus yields better results (in terms of higher BLEU scores), whereas for translating into English it is better to use the former.

Machine translation implements natural language processing, a task of generating natural language from a machine representation system such as a knowledge base or a logical form; Psycholinguists prefer the term language production when such formal representations are interpreted as models for mental representations.

It could be said an NLG system is like a translator that converts a computer based representation into a natural language representation. However, the methods to produce the final language are different from those of a compiler due to the inherent expressivity of natural languages.

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:6(1), June 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

NLG may be viewed as the opposite of natural language understanding: whereas in natural language understanding the system needs to disambiguate the input sentence to produce the machine representation language, in NLG the system needs to make decisions about how to put a concept into words. Simple examples are systems that generate form letters. These do not typically involve grammar rules, but may generate a letter to a consumer, e.g. stating that a credit card spending limit was reached. More complex NLG systems dynamically create texts to meet a communicative goal. As in other areas of natural language processing, this can be done using either explicit models of language (e.g., grammars) and the domain, or using statistical models derived by analysing human-written texts.

Despite the fact that image understanding and natural language processing constitute two major areas, there have only been a few attempts toward the integration of computer vision and the generation of natural language expressions for the description of image sequences. Although there has been much progress in developing theories, models and systems in the areas of Natural Language Processing (NLP) and Vision Processing (VP) there has been little progress on integrating these two subareas of Artificial Intelligence (AI). In the beginning the general aim of the field was to build integrated language and vision systems, few were done, and two subfields quickly arose. It is not clear why there has not already been much activity in integrating NLP and VP. Is it because of the long-time reductionist trend in science up until the recent emphasis on chaos theory, non-linear systems, and emergent behaviour? Or, is it because the people who have tended to work on NLP tend to be in other Departments, or of a different ilk, to those who have worked only in VP?

There has been a recent trend towards the integration of NLP and VP and other forms of perception such as speech. Dennett (1991, pp. 57-58) says "Surely a major source of the widespread skepticism about "machine understanding" of natural language is that such systems almost never avail themselves of anything like a visual workspace in which to parse or analyze the input. If they did, the sense that they actually understood what they processed would be greatly heightened (whether or not it would still be, as some insist, an illusion). As it is, if a computer says, "I see what you mean" in response to input, there is a strong temptation to dismiss the assertion as an obvious fraud." Heretofore social trends in science in general have been towards reductionism. Pure reductionism argues that the social nature of experimentation is irrelevant to scientific outcome where the interactions between scientists should have no effect upon their results. In contrast, the effort on integration required here will certainly involve social interaction between researchers in each field which might not have occurred otherwise. What matters to scientific progress is not the conducting of experiments per se, but rather the determination of which experiments are worth conducting. In such contexts, 'worth' is clearly a sociological, as opposed to a scientific, matter.

Computational generation of referring expressions in visual translation:

Consider three balls, suppose one wants to point out a ball. Most speakers have no difficulty in accomplishing this task, by producing a referring expression such as “the red ball” for example. Now imagine a computer being confronted with the same task, aiming to point out the ball d2. Assuming it has access to a database containing all the relevant properties of the balls in the scene, it needs to find some combination of properties which applies to d2, and not to the other two. There is a choice though: There are many ways in which d2 can be set apart from the rest (“the ball in the middle” “the red ball” “the ball to the right of the blue ball”), and the computer has to decide which of these is optimal in the given context. Moreover, optimality can mean different things. It might be thought, for instance, that references are optimal when they are minimal in length, containing just enough

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:6(1), June 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

information to single out the target. But, as we shall see, finding minimal references is computationally expensive, and it is not necessarily what speakers do, nor what is most useful to hearers.

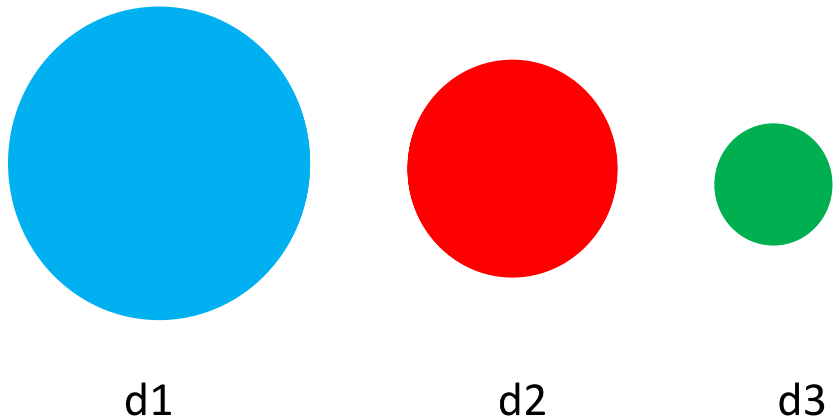

Figure 1: A simple visual scene

Referring expressions play a central role in communication, and have been studied extensively in many branches of (computational) linguistics, including Natural Language Generation (NLG). NLG is concerned with the process of automatically converting non-linguistic information into natural language text, which is useful for practical applications ranging from generating weather forecasts to summarizing medical information. Of all the subtasks of NLG, Referring Expression Generation (REG) is among those that have received most scholarly attention. A survey of implemented, practical NLG systems shows that virtually all of them, regardless of their purpose, contain an REG module of some sort. This is hardly surprising in view of the central role that reference plays in communication. A system providing advice about air travel (White, Clark, and Moore 2010) needs to refer to flights ("the cheapest flight," "the KLM direct flight"), a pollen forecast system (Turner et al. 2008) needs to generate spatial descriptions for areas with low or high pollen levels ("the central belt and further North"), and a robot dialogue system that assembles construction toys together with a human user (Giuliani et al. 2010) needs to refer to the components ("insert the green bolt through the end of this red cube"). REG "is concerned with how we produce a description of an entity that enables the hearer to identify that entity in a given context" (Reiter and Dale 2000, page 55). Because this can often be done in many different ways, a REG algorithm needs to make a number of choices. According to Reiter and Dale (2000), the first choice concerns what *form* of referring expression is to be used; should the target be referred to, for instance, using its proper name, a pronoun ("he"), or a description ("the man with the tie"). Proper names have limited applicability because many domain objects do not have a name that is in common usage. For pronoun generation, a simple but conservative rule is discussed by Reiter and Dale (2000), similar to one proposed by Dale (1989, pages 150–151): Use a pronoun if the target was mentioned in the previous sentence, and if this sentence contained no reference to any other entity of the same gender. Reiter and Dale (2000) concentrate mostly on the generation of descriptions. If the NLG system decides to generate a description, two choices need to be made: Which set of properties distinguishes the target (content selection), and how the selected properties are to be turned into natural language (linguistic realization). One of the most ubiquitous tasks in natural language generation is the generation of referring expressions: phrases that identify particular domain entities to the human recipient of the generation system's output. Each property expressed in a referring expression can be regarded as having the function of 'ruling out' members of the contrast

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:6(1), June 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

set. Suppose a speaker wants to identify a small black dog in a situation where the contrast set consists of a large white dog and a small black cat. She might choose the adjective black in order to rule out the white dog and the head noun dog in order to rule out the cat; this would result in the generation of the referring expression the black dog, which matches the intended referent but no other object in the current context. The small dog would also be a successful referring expression in this context, under the distinguishing description model. More formally, we assume that each entity in the domain is characterised by means of a set of attribute–value pairs. We will sometimes to refer to an attribute–value pair as a property. We assume that the semantic content of a referring expression can also be represented as a set of attribute–value pairs. We will use the notation (Attribute, Value) for attribute–value pairs; for example, (colour, red) indicates the attribute of colour with the value red. The semantic content of head nouns will be represented as the value of the special attribute type: for example, (type, dog). Let r be the intended referent, and C be the contrast set; then, a set L of attribute–value pairs will represent a distinguishing description if the two conditions in hold:

- Every attribute–value pair in L applies to r: that is, every element of L specifies an attribute–value that r possesses.
- For every member c of C, there is at least one element l of L that does not apply to c: that is, there is an l in L that specifies an attribute–value that c does not possess. l is said to rule out c.

For example, suppose the task is to create a referring expression for Object1 in a context that also includes Object2 and Object3, where these objects possess the following properties:

• Object1: <type, dog>, <size, small>, <colour, black>

• Object2: <type, dog>, <size, large>, <colour, white>

• Object3: <type, cat>, <size, small>, <colour, black>

In this situation, r = Object1 and C = {Object2, Object3}. The content of one possible distinguishing description is then {<type, dog>, <colour, black>}, which might be realised as the black dog: Object1 possesses these properties, but Object2 and Object3 do not (Object2 does not have the property <colour, black>, while Object3 is ruled out by <type, dog>).

A good referring expression generation algorithm should be able to take into account what is known about the hearers/ reader's knowledge and perceptual abilities. This can be done at the simplest level by restricting the attributes mentioned in a referring expression to those which most human hearers/reader's are presumed to easily be able to perceive, such as colour and size. A more general solution is to allow the generation algorithm to issue appropriate queries to a hearer/reader's model at run-time. There are many criteria that an algorithm which generates referring expressions should satisfy; we are particularly concerned here with the following:

1. The algorithm should generate referring expressions which satisfy the referential communicative goal: after hearing or seeing the referring expression, the human hearer or reader should be able to identify the target object.

2. The algorithm should generate referring expressions which do not lead the human hearer or reader to make false conversational interpretations.

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:6(1), June 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

3. The algorithm, if it is to be of practical use, should be computationally efficient.

Computational journalism: as an application of visual translation:

Computational journalism combines classic journalistic values of storytelling and public accountability with techniques from computer science, statistics, the social sciences, and the digital humanities. It can be defined as the application of computation to the activities of journalism such as information gathering, organization, sense making, communication and dissemination of news information, while upholding values of journalism such as accuracy and verifiability. The field draws on technical aspects of computer science including artificial intelligence, content analysis (NLP, vision, audition), visualization, personalization and recommender systems as well as aspects of social computing and information science. Computational reporting can be evolved over time to inform visual news such as that of reporting accidents etc., as well as discrete data to the mass audience. For example the pollen forecast for Scotland, shows a simple NLG system in action. This system takes as input six numbers, which give predicted pollen levels in different parts of Scotland. From these numbers, the system generates a short textual summary of pollen levels as its output. For example, using the historical data for 1-July-2005, the software produces,

Grass pollen levels for Friday have increased from the moderate to high levels of yesterday with values of around 6 to 7 across most parts of the country. However, in Northern areas, pollen levels will be moderate with values of 4.

In contrast, the actual forecast (written by a human meteorologist) from this data was

Pollen counts are expected to remain high at level 6 over most of Scotland, and even level 7 in the south east. The only relief is in the Northern Isles and far northeast of mainland Scotland with medium levels of pollen count.

The process to generate reports can be as simple as keeping a list of canned text that is copied and pasted, possibly linked with some glue text. The results may be satisfactory in simple domains such as horoscope machines or generators of personalised business letters. However, a sophisticated NLG system needs to include stages of planning and merging of information to enable the generation of text that looks natural and does not become repetitive. The typical stages of natural language generation, as proposed by Dale and Reiter are

- ✓ Content determination: selecting the information to be mentioned in the text.
- ✓ Document structuring: organisation for information conveying
- ✓ Aggregation: To improve naturalness of the report.
- ✓ Lexical choice: Putting appropriate words to the context.
- ✓ Referring expression generation: Creating expressions for identifying objects and regions
- ✓ Realisation: Creating correct text obeying the rules of syntax, morphology, and orthography.

Conclusion:

Visual translation not only erases the barricades of language diversity but also the gap between machine and man communication. We believe that it is time for evaluation studies to extend their remit and look at the types of complex references that more recent REG research has drawn attention to. Such studies would do well, in our view, to pay considerable attention to the question of which referring expressions have the greatest benefit for readers or hearers. One day, perhaps, all these issues will have been resolved. If there is anything that a survey of the state of the art in REG makes clear it is that, for all the undeniable progress in this growing area of NLG, this Holy Grail is not yet within reach.

References:

1. Belz, Anja. 2009. That's nice . . . what can you do with it? [Last Words].Computational Linguistics, 35:111–118.
2. Appelt, Douglas. 1985. Planning English referring expressions. Artificial Intelligence, 26:1–33.
3. Gatt, Albert and Anja Belz. 2010. Introducing shared task evaluation to NLG: The TUNA shared task evaluation challenges. In Emiel Krahmer and Mariet Theune, editors, Empirical Methods in Natural Language Generation; Springer Verlag, Berlin, pages 264–293.
4. Horacek, Helmut. 1997. An algorithm for generating referential descriptions with flexible interfaces. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL), pages 206–213, Madrid.
5. Paraboni, Ivandre, Kees van Deemter, ´ and Judith Masthoff. 2007. Generating referring expressions: Making referents easy to identity. Computational Linguistics, 33:229–254.
6. Appelt, D & Kronfeld, A (1987); A computational model of referring; In Proceedings of IJCAI-1987, volume 2, pages 640–647.
7. Dale, Robert; Reiter, Ehud (2000); *Building natural language generation systems*. Cambridge, U.K.: Cambridge University Press.
8. Goldberg E, Driedger N, Kittredge R (1994). "Using Natural-Language Processing to Produce Weather Forecasts"