Paraphrase Concept and Typology.
A Linguistically Based and Computationally Oriented Approach[*]

Concepto y tipología de paráfrasis.
Una aproximación lingüística orientada al tratamiento computacional

Marta Vila, M. Antònia Martí
CLiC-UB
Gran Via 585, 08007 Barcelona
marta.vila@ub.edu, amarti@ub.edu

Horacio Rodríguez
TALP-UPC
Jordi Girona 1-3, 08034 Barcelona
horacio@lsi.upc.es

Resumen: En este artículo, se presenta un análisis crítico de la bibliografía sobre la definición de paráfrasis y su tipología. Dicho análisis pone de manifiesto que no existe una caracterización de la paráfrasis completa y lingüísticamente fundamentada que, al mismo tiempo, sea tratable computacionalmente. Se propone una definición y delimitación del concepto fundada sobre el contenido proposicional. Sobre esta base, se ha elaborado una tipología general, inclusiva y orientada al tratamiento computacional de los mecanismos lingüísticos que dan lugar a la variación en la forma de los pares parafrásticos.
Palabras clave: Paráfrasis, límites de la paráfrasis, tipología de paráfrasis.

Abstract: In this paper, we present a critical analysis of the state of the art in the definition and typologies of paraphrasing. This analysis shows that there exists no characterization of paraphrasing that is comprehensive, linguistically based and computationally tractable at the same time. The following sets out to define and delimit the concept on the basis of the propositional content. We present a general, inclusive and computationally oriented typology of the linguistic mechanisms that give rise to form variations between paraphrase pairs.
Keywords: Paraphrasing, paraphrase boundaries, paraphrase typology.

1 Introduction

Paraphrasing stands for sameness of meaning between different wordings. Prototypical paraphrase examples can be seen in (1) and (2), where the semantic content remains the same despite the differences in the form: *significant* is substituted for its synonym *considerable* in (1-b), and (2) illustrates an active/passive diathesis alternation.

(1)   a.   This task requires *significant* knowledge to be successful.
      b.   This task requires *considerable* knowledge to be successful.

(2)   a.   The Romans constructed that bridge.
      b.   That bridge was constructed by the Romans.

The omnipresence of paraphrasing in natural language gives rise to the need to apprehend the mechanisms that govern this phenomenon from a linguistic perspective. Natural Language Processing (NLP) components dealing with paraphrasing, in turn, appear to have great potential for the improvement of systems for understanding and generation, such as question answering, summarization or machine translation. Despite its potential, a linguistically backed and, at the same time, computationally efficient account of the whole paraphrase phenomenon has not yet been developed.

In this work, a proposal for the characterization of paraphrasing is presented. We follow two different perspectives: an intensional perspective setting out the properties a linguistic expression needs to be considered a paraphrase (the concept), and an extensional perspective specifying the objects that fall under paraphrasing (typology). It consists in a comprehensive and linguistically founded

approach with a practical orientation regarding its potential computational applications.

In the following, we introduce the state of the art in paraphrase characterization (Section 2). In Section 3, our paraphrase typology is presented, backed by a concept definition. This section is illustrated with examples extracted/adapted from different sources in Section 2 and our own examples. Sections 4 and 5 set out the conclusions and future work, respectively.

2   State of the Art

As our work tries to approach the phenomenon with linguistic rigor while making it computationally treatable, we present what has been said about the paraphrase concept (Section 2.1) and typology (Section 2.2) both in Linguistics and NLP.

2.1   The Concept

In Linguistics, paraphrasing is addressed both from discourse theory and from linguistic analysis.

In **discourse analysis**, works focus on the reformulation mechanisms or the communicative intention behind them (Gülich and Kotschi, 1995), without considering the linguistic nature of paraphrases themselves. For this reason, we are not going to go into this group in depth.

From the perspective of **linguistic analysis**, as paraphrasing consists of different forms expressing the same content, it is placed in the syntax-semantic interface. This issue has been addressed by different theories.

Meaning-Text Theory (MTT) (Mel'čuk, 1992) understands paraphrases as those linguistic expressions ("Text") sharing the same semantic content ("Meaning"). More specifically, departing from an underlying representation, all corresponding sentences—paraphrases between them—are derived without modifying the underlying representation itself (meaning correspondence); paraphrases can also be obtained from reformulated underlying representations (meaning equivalence). However, the axiomatic foundations of MTT prevent the exploitation of the paraphrase system outside its own framework.

Transformations and diathesis alternations can be considered indirect approaches to paraphrasing in the sense that they deal with equivalent expressions obtained by transformations between superficial structures (Harris, 1957), transformation from deep to superficial representations (Chomsky, 1957) and the different realization of predicates (Levin, 1993). Indeed, some authors have explicitly taken advantage of them to explain paraphrasing (Culicover, 1968; Fujita, 2010). The fact that these theories have been explored by many authors and received varied reformulations, have made them part of the general linguistic background. However, in paraphrase terms, the preservation of the meaning in these transformations is controversial. Moreover, they do not cover paraphrasing as a whole, but focus on syntactic phenomena.

In **NLP**, although paraphrasing is part of many tasks, systems and applications,[1] works have hardly delved into the linguistic nature of the phenomenon. They rely on very simplistic and vague definitions, on many occasions, quantitative in nature (e.g., number of common words or n-grams). However, some considerations should be set out.

From the field of logic, Dras (1999) sets out a definition of paraphrasing based on the model of truth-conditional semantics: two units of text are interchangeable if, for the propositions A and B they embody, the truth-set of B is a (not necessarily proper) subset of the truth-set of A. Authors that work in textual entailment, in turn, define paraphrasing as bidirectional entailment (Rus et al., 2009). These definitions do not overcome the difficulty of establishing sameness of meaning but transfer it to other domains: the establishment of the truth-set of the propositions and the textual entailment recognition. Moreover, their rigidness omits many phenomena that should fall under paraphrasing.

In contrast to the more restrictive logic based definitions of paraphrasing, Bhagat (2009) states that, although some potential paraphrases are not equivalent in the logical sense, they must be considered paraphrases (or 'quasi-paraphrases') for all practical purposes. Thus, authors working in paraphrasing generally qualify paraphrases as having *approximate* semantic equivalence. This definition, being more inclusive than the logic based ones, presents the problem of the vagueness in terms of paraphrase boundaries.

---

[1]See surveys by Androutsopoulos and Malakasiotis (2010), and Madnani and Dorr (2010).

2.2 Typologies

Analogously to Section 2.1, we classify existing paraphrase typologies into 3 groups: typologies coming from discourse analysis, linguistic analysis and NLP.

Typologies coming from **discourse analysis** classify paraphrases according to the reformulation mechanisms or communicative intention behind them, e.g., modification, exemplification, metaphor and simile, and arguing (Mei Ling Lisa, 2009). They shed light on several discourse-related issues in paraphrasing, but their purposes are different from ours.

Few paraphrase typologies have been built from **linguistic analysis** and some of them are rigidly tied to the framework of a theory. MTT defines a list of lexical paraphrasing rules expressed in terms of lexical functions; along with a list of syntactic paraphrasing rules expressed in terms of dependency trees indicating what restructuring of the deep syntactic level is needed when a particular lexical rule is applied (Mel'čuk, 1992). Milićević (2007), in turn, describes a typology operating at the semantic level in terms of propositional and communicative equivalences. A drawback presenting these paraphrase systems is that they are hardly generalizable outside the MTT framework.

Some typologies are less tied to concrete theoretical frameworks, but, at the same time, are less formal and computationally treatable. Apresjan (1973) mainly deals with lexical paraphrases. Martin (1976), in turn, presents a general account of the phenomenon, but his approach focuses mainly on connotation, opposition and synonymy based paraphrases.

There exist a number of paraphrase typologies build in **NLP** works. Some of these typologies are simply lists of paraphrase types useful for a specific system or application, or the most common types found in a corpus. They are specific-work oriented and far from being comprehensive. Some examples of these typologies are Barzilay, McKeown, and Elhadad (1999), Kozlowski, McCoy, and Shanker (2003) and Dorr et al. (2004).

Other typologies classify paraphrasing in a very generic way, only setting out two or three types. They would not reach the category of typologies *strictu sensu*. Some examples of these typologies are Barzilay (2003) and Shimohata (2004).

Finally, there are more comprehensive typologies like the ones by Culicover (1968), Dras (1999), Fujita (2005), and Bhagat (2009). They usually take the shape of extensive and very fine-grained lists of paraphrasing types grouped into bigger classes following different criteria. These approaches are the most closely related to what we are presenting here. That is why we are looking into them in more detail.

Although it is an early work, Culicover (1968) displays good intuition in many respects. First, he sets out a linguistically founded and coherent grouping of the paraphrasing types. Secondly, he carries out a formalization attempt through the definition of some structural and semantic conditions to be fullfilled by each of the paraphrase types.

A more recent work, Dras (1999), carries out a deep and rigorous study of syntactic paraphrases as a basis for their formal representation using Synchronous Tree Adjoining Grammars (STAG). While not being the focus of Dras (1999), the general classes in which the paraphrasing types are grouped are created following different criteria (according to the formal change observed in the paraphrase pair, e.g., change of order, or according to the paraphrase effect, e.g., change of focus), which makes them not mutually exclusive (a change of order and a change of focus can coexist).

Fujita (2005; 2010) are the most comprehensive works. They build the most general and exhaustive paraphrase typology including paraphrases at all language levels. Despite being very different from Dras (1999) in nature, the classes grouping the paraphrase types also lack a clear criteria behind them (e.g., extra-sentential, extra-clausal, pure syntactic and lexical paraphrases).

Finally, Bhagat (2009) classifies all paraphrases according to their lexical basis (e.g., actor/action substitution or noun/adjective conversion) and links each of these types to the structural modifications involved (substitution, addition/deletion and/or permutation). Although this is a successful approach looking for a simple and clear explanation of the paraphrase phenomenon, it presents the problem of excessive simplification, especially in limiting the possible structural modifications to only three.

A final consideration regarding all NLP paraphrase typologies should be set out.

They generally focus on the specific paraphrase mechanisms, leaving the general phenomena at a second level. However, only the latter can account for paraphrasing in a comprehensive way. A list of specific mechanisms will always be endless.

3 Paraphrase Characterization

Our objective is building a linguistically based and, at the same time, computationally oriented characterization of paraphrasing. We understand a computationally oriented paraphrase characterization as the description of the subset of paraphrase phenomena that are treatable with state-of-the-art computational resources. Paraphrases that cannot be addressed computationally are beyond the scope of our work.

In Section 3.2, we present our proposal for a paraphrase typology. In Section 3.1, the concept of paraphrasing laying behind this typology is defined.

3.1 The Concept

Paraphrasing is a complex phenomenon whose boundaries with other linguistic facts, such as coreference, inference or pragmatic relationships are sometimes blurred.

Paraphrasing and coreference overlap considerably, but their definitions make them notably different in essence: paraphrasing concerns meaning, whereas coreference is about discourse referents. In this sense, and in accordance with the approach by Fuchs (1994), we consider (3)-like paraphrases that Fujita (2005) and Milićević (2007) call, respectively, "referential" and "cognitive" to be best treated as coreference rather than paraphrase, because they only rely on referential identity in a discourse, not on sameness of meaning (Recasens and Vila, 2010).

(3)   a.   They got married *last year.*
       b.   They got married *in 2004.*

Similarly, Dorr et al. (2004)'s "viewpoint variation paraphrases" are not considered here either. Sentences in (4) do no have the same meaning, although they can have the same referent.

(4)   a.   The U.S.-led *invasion* of Iraq
       b.   The U.S.-led *liberation* of Iraq

Although their boundaries are not clear either, paraphrasing and inference stand for significantly different phenomena: while the former is about sameness of meaning, which is a bidirectional quality, the latter consists of unidirectional entailment relationships. We consider that (5)-like "semantic implication" and "inference", treated as paraphrase types by Bhagat (2009) and Dorr et al. (2004) respectively, should be addressed in the field of textual entailment.

(5)   a.   Google *was in talks to buy* YouTube
       b.   Google *bought* YouTube

We also consider Fujita (2005)'s "pragmatic paraphrases" exemplified in (6) to be outside the limits of paraphrasing, as they do not hold the same meaning, but the same ilocutive value.

(6)   a.   I want some fresh air.
       b.   Could you open the window?

The examples above illustrate where the boundaries between paraphrasing and paraphrase related phenomena should be drawn. In the following, the nature of what is within paraphrase boundaries is set out. In our approach, the propositional content, i.e., what is explicitly expressed, and the form are taken as the point of departure for defining paraphrasing.

We hypothesize that there exists a correlation between the differences in propositional content and the differences in wording on the one hand, and the degree of sameness of meaning or paraphrasability on the other, both being gradual properties: i) paraphrases with the same propositional content and very similar in form exhibit a high degree of paraphrasability; ii) in paraphrases with equivalent propositional content and a slightly different form, the degree of paraphrasability is lower; iii) in paraphrases with equivalent propositional content but a completely different wording, the paraphrasability decreases even more.

The paraphrase pair in (7) exhibits the same propositional content as only a change in the PoS has occurred: *reduction* is changed to its verbal form *reduced.* Thus, a formal mapping between the two members of the pair can be established straightforwardly. The degree of paraphrasability is high.

The paraphrase pair in (8) does not exhibit the same, but rather an equiva-

lent propositional content, because it consists in a change in the lexicalization pattern: the meaning components are the same (MOVEMENT, MEANS and PATH), but they are lexicalized differently (units in italis, in bold or underlined). Although more different in form than the pair in Example (7), a formal mapping is still possible. The degree of paraphrasability is, in that case, lower.

The paraphrase pair in (9) cannot be formally mapped. The degree of paraphrasability is even lower, although still showing an equivalent propositional content.

(7)   a.   The tenants wanted a *reduction* in the charge for electricity

        b.   The tenants wanted the charge for electricity to be *reduced*

(8)   a.   Jim *flew* **across** the <u>sea</u>

        b.   Jim **crossed** the <u>ocean</u> *by plane*

(9)   a.   She is thorough in everything she does

        b.   Always taking care of every detail is one of her characteristics

Paraphrases in iii), i.e., (9)-like examples, are not addressed in our typology, as their lack of formal mapping prevents its computational treatment.

3.2 Typology

The typology we are presenting here (Table 1) classifies paraphrases according to the linguistic nature of their difference in wording. It attempts to capture the general linguistic phenomena of paraphrasing, rather than presenting a long, coarse-grained and inevitably incomplete list of concrete mechanisms.

Our typology consists of a two-level typology of 9 paraphrasing types grouped into 5 classes. Paraphrasing types reflect a general paraphrase mechanism (e.g., deletion or diathesis alternation); their grouping, in turn, reflects the level of language where this mechanism takes place (e.g., lexicon or syntax based changes, respectively).

Belonging to one group does not necessarily mean that the level of language affected by the change is only that which gives name to the group, but that the change appears at that level and other levels can be affected: a lexicon motivated change like the one in (10-a)–(10-b), where *said* is changed for its synonym *told*, has syntactic implications, in this case, a change in the syntactic realiza-

tion of the arguments. Moreover, although types in our typology are presented in isolation, they can be combined: in (10-c)–(10-a) both a same-polarity lexical substitution and a change in the discourse structure can be observed.

(10)   a.   Bill *said* to me that she was Irish

         b.   Bill *told* me that she was Irish

         c.   What Bill told me is that she was Irish

Our typology fits our idea of the paraphrase concept. Morphology, syntax and discourse based changes mantain the same propositional content. Lexicon and semantics based ones, in turn, are equivalent in their propositional content. It has to be said that, in discourse based changes, the change in the wording can be more important, but it only changes in meaningless units.

The five classes in which our typology is divided comprise lexicon, morphology, syntax, semantics and discourse based changes.

Lexicon based changes consist in changing one lexical unit for another one. It comprises same and opposite polarity substitutions, deletions and synthetic/analytic substitutions.

<u>Same polarity substitutions</u> change one lexical unit for another one with the same meaning. Synonymy substitutions (Example (1) in Table 1), exact/approximate substitutions (2) or inductor/induced agent substitutions (3) are examples of this paraphrase type.

<u>Opposite polarity substitutions</u> consist in changing one lexical unit for another one with opposite polarity (antonyms, complementaries or conversives). In order to maintain the same meaning, an argument inversion or a double change of polarity have to occur. In (4), an antonymy substitution (*leaving/staying*) plus a negation occur. In (5), the change to a complementary lexical unit (*Only 20%/Most*) is compensated with the antonymy (*on time/late*). In (6), an argument inversion and a conversive substitution (*bought/was sold to*) take place.[2]

<u>Deletions</u>[3] consist in the substitution of

---

[2]Martin (1976) sets out a complete characterization of the "double negation and double inversion paraphrase", the basis of opposite polarity substitutions in our typology.

[3]Some paraphrasing types can be seen from two opposite perspectives, e.g., deletion/addition. We

one text snippet for the vacuum. Deleted snippets can be lexical, or non- or semi-lexical. In (7), the verb *eating* has been deleted. In (8), it is a semi-lexical unit (*actually*) which has been deleted.

Finally, synthetic/analytic substitutions are those substitutions of single-pieced for multiple-pieced lexical units with the same meaning. This type comprises cases such as light element addition (*made* in (9-b)), specifier addition (*sequence* in (10-b)) or compounding/decomposition (11).

Morphology based changes are those that arise at the morphology level of language. They can be inflectional or derivational.

Inflectional changes consist in adding inflectional affixes to words. An example of this type can be found in (12), where the meaning of the past (12-a) and the so called historical present (12-b) is the same.

Derivational changes consist in the addition of derivational affixes to words. In (13), the addition of the suffix *-ation* causes a change in the PoS of the verb *founded* and converts it into a noun, with the consequent reorganization of the sentence structure.

Syntax based changes are about syntactic reorganization and contain those diathesis alternations in which there is meaning maintenance, such as active/passive (14), transitivity (15) and locative alternations (16).

Semantics based changes are those that imply a different lexicalization pattern for the same content units. In (17), (18) and (19), the lexical units in bold and in italics express the same semantic content, respectively.

Finally, **discourse based changes** are those changing the discourse structure of the sentence. This group covers a broad range of discursive reorganizations, including phenomena such as functional word changes (20), relative clause deletion (21) and sentence splitting/combining (22).

4   Conclusions

In this paper, we have presented a linguistically based and computationally oriented paraphrase typology backed by a definition and delimitation of the phenomenon. A paraphrase typology constitutes a key tool for

---

only refer to one of them (deletion), leaving the other one (addition) as understood.

many paraphrase related NLP modules and applications, e.g., it paves the way for determining which paraphrasing types a system fails to address, or which would be the best way to solve the problem according to the linguistic nature of these types. In that sense, we respond to the call by Wintner (2009), who demonstrates what NLP can achieve when it is backed by linguistic grounds.

5   Future Work

A validation of our typology using it to manually classify paraphrases extracted by the WRPA method (Vila, Rodríguez, and Martí, 2010) and paraphases in the PAN-PC-10 corpus for plagiarism detection (Potthast et al., 2010) is currently under development. The annotation of plagiarism examples following our typology will shed light to the potential benefits of using paraphrase knowledge and techniques in plagiarism detection. Finally, we are working on the formalization of the typology trying to see which would be the best approach(es) to carry it out and to what extend paraphrase types are fully formalizable.

References

Androutsopoulos, Ion and Prodromos Malakasiotis. 2010. A survey of paraphrasing and textual entailment methods. *Journal of Artificial Intelligence Research*, 38:135–187.

Apresjan, Jurij Derenikowicz. 1973. Synonymy and synonyms. In Ferenc Kiefer, editor, *Trends in Soviet Theoretical Linguistics*. D. Reidel Publishing Company, pages 173–199.

Barzilay, Regina. 2003. *Information Fusion for Multidocument Summarization: Paraphrasing and Generation*. Ph.D. thesis, Columbia University.

Barzilay, Regina, Kathleen McKeown, and Michael Elhadad. 1999. Information fusion in the context of multi-document summarization. In *Proceedings of the ACL 1999*, pages 550–557.

Bhagat, Rahul. 2009. *Learning Paraphrases from Text*. Ph.D. thesis, University of Southern California.

Chomsky, Noam. 1957. *Syntactic Structures*. Mouton & Co.

Culicover, Peter. 1968. Paraphrase generation and information retrieval from stored text. *Mechanical Translation and Computational Linguistics*, 11(1 and 2):78–88.

Dorr, Bonnie J., Rebecca Green, Lori Levin, Owen Rambow, David Farwell, Nizar Habash, Stephen Helmreich, Eduard Hovy, Keith J. Miller, Teruko Mitamura, Florence Reeder, and Advaith Siddharthan. 2004. Semantic annotation and lexico-syntactic paraphrase. In *Proceedings of the Workshop on Building Lexical Resources from Semantically Annotated Corpora, LREC 2004*.

Dras, Mark. 1999. *Tree Adjoining Grammar and the Reluctant Paraphrasing of Text.* Ph.D. thesis, Macquarie University.

Fuchs, Catherine. 1994. *Paraphrase et énonciation.* Ophrys.

Fujita, Atsushi. 2005. *Automatic Generation of Syntactically Well-formed and Semantically Appropriate Paraphrases.* Ph.D. thesis, Nara Institute of Science and Technology.

Fujita, Atsushi. 2010. Typology of paraphrases and approaches to compute them. Invited talk to the CBA 2010 workshop: Corpus-based Approaches to Paraphrasing and Nominalization, Barcelona.

Gülich, Elizabeth and Thomas Kotschi. 1995. Discourse production in oral communication. In Quasthoff Uta M., editor, *Aspects of Oral Comunication. Research in Text Theory.* Walter de Gruyter, pages 30–66.

Harris, Zellig. 1957. Co-occurence and transformation in linguistic structure. *Language*, 3(33):283–340.

Kozlowski, Raymond, Kathleen F. McCoy, and Vijay K. Shanker. 2003. Generation of single-sentence paraphrases from predicate/argument structure using lexico-grammatical resources. In *Proceedings of IWP 2003*, pages 1–8.

Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation.* University of Chicago Press.

Madnani, Nitin and Bonnie J. Dorr. 2010. Generating phrasal and sentential paraphrases: A survey of data-driven methods.

Computational Linguistics, 36(3):341–387.

Martin, Robert. 1976. *Inférence, antonymie et paraphrase.* Librarie C. Klincksieck.

Mei Ling Lisa, Cheung. 2009. *Merging Corpus Linguistics ans Collaborative Knowledge Construction.* Ph.D. thesis, University of Birmingham.

Mel'čuk, Igor A. 1992. Paraphrase et lexique: la théorie sens-texte et le *Dictionnaire explicatif et combinatoire.* In Igor A. Mel'čuk et al., editors, *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques III.* Les Presses de l'Université de Montréal, pages 9–59.

Milićević, Jasmina. 2007. *La paraphrase.* Peter Lang, Berne.

Potthast, Martin, Benno Stein, Alberto Barrón-Cedeño, and Paolo Rosso. 2010. An evaluation framework for plagiarism detection. In *Proceedings of the COLING 2010*, pages 997–1005.

Recasens, Marta and Marta Vila. 2010. On paraphrase and coreference. *Computational Linguistics*, 36(4):639–647.

Rus, Vasile, Philip M. McCarthy, Arthur C. Graesser, and S. McNamara Danielle. 2009. Identification of sentence-to-sentence relations using a textual entailer. *Research on Language and Computation*, 7(2–4):209–229.

Shimohata, Mitsuo. 2004. *Acquiring Paraphrases from Corpora and Acquiring Paraphrases from Corpora and Its Application to Machine Translation.* Ph.D. thesis, Nara Institute of Science and Technology.

Vila, Marta, Horacio Rodríguez, and M. Antònia Martí. 2010. WRPA: A system for relational paraphrase acquisition from Wikipedia. *Procesamiento del Lenguaje Natural*, 45:11–19.

Wintner, Shuly. 2009. What science underlies Natural Language Engineering? *Computational Linguistics*, 35(4):641–644.

Lexicon based changes			
Same polarity substitution	(1)	a.	Google *bought* YouTube
		b.	Google *acquired* YouTube
	(2)	a.	They were *9*
		b.	They were *around 10*
	(3)	a.	The *pilot* took off despite the stormy weather
		b.	The *plane* took off despite the stormy weather
Opposite polarity substitution	(4)	a.	I am *leaving*
		b.	I am *not staying*
	(5)	a.	*Only 20%* of the participants arrived *on time*
		b.	*Most* of the participants arrived *late*
	(6)	a.	Google *bought* YouTube
		b.	YouTube *was sold to* Google
Deletion	(7)	a.	I like *eating* chocolate
		b.	I like chocolate
	(8)	a.	*Actually,* you shouldn't be here
		b.	You shouldn't be here
Synthetic/analytic substitution	(9)	a.	Steven *attempted* to stop playing Hearts
		b.	Steven *made an attempt* to stop playing Hearts
	(10)	a.	*Ideas* is all I need to write an article
		b.	*A sequence of ideas* is all I need to write an article
	(11)	a.	I prefer *wildlife television documentaries*
		b.	I prefer *television documentaries about wildlife*
Morphology based changes			
Inflectional change	(12)	a.	In 1492 Columbus *reached* America
		b.	In 1492 Columbus *reaches* America
Derivational change	(13)	a.	I know that Olds *founded* GM
		b.	I know about the *foundation* of GM by Olds
Syntax based changes			
Diathesis alternation	(14)	a.	John loves Mary
		b.	Mary is loved by John
	(15)	a.	The laundry sways in the breeze
		b.	The breeze makes the laundry sways
	(16)	a.	The section chief filled Japanese sake into the cup
		b.	The section chief filled the cup with Japanese sake
Semantics based changes			
Change in the lexicalization pattern	(17)	a.	Bill *flew* **across** the ocean
		b.	Bill **crossed** the ocean *by plane*
	(18)	a.	The *increase* of prices **accompanies** the crise
		b.	The prices *increasent* **with** the crise
	(19)	a.	Barbara *excels* at **teaching**
		b.	Barbara **teaches** *well*
Discourse based changes			
Change in the discourse structure	(20)	a.	He wanted to eat *nothing but* apples
		b.	*All* he wanted to eat *were* apples
	(21)	a.	Joe wants the blazer *which was* designed by BMW
		b.	Joe wants the blazer designed by BMW
	(22)	a.	He is willing to leave. This made Gillian upset
		b.	His willingness to leave made Gillian upset

Table 1: Paraphrase typology