The International Conference on Interdisciplinary Humanities and Communication Studies
DOI: 10.54254/2753-7064/3/20220237

A Typological Approach to Imperfect: Encoding Types and Semantic Features

Xiuyi Wu[1,a,*], Haoran Ma[1]

[1]*College of Foreign Studies, Jinan University, Guangzhou, 511400, China*
a. laylawu@stu2019.jnu.edu.cn
**corresponding author*

Abstract: Imperfect was first referred to in the linguistic field quite a long time ago as a special combination of tense and aspect. However, it has long been regarded as subordinate to Imperfective and to exist merely in some Romance, Latin, Greek, Indo-Iranian and Balto-Slavonic languages, thus receiving little specialized research. To conduct a thorough and detailed analysis of Imperfect, the present study established 105 language samples across the world. The imperfect meaning, with different encoding devices, is found in most language samples around the world. Based on the typological classification of their encoding types, the semantic features of Imperfect are then discussed to provide a better comprehension of the universals and characteristics of the types. This multi-dimensional research attempts to call attention to this underestimated concept in linguistic research into tense and aspect, and to provide the research of Imperfect with a rationale in expectation of more systematic studies of Imperfect in the future.

Keywords: Imperfect, Typology, Encoding Types, Semantic Features

1. Introduction

Etymologically, the word “imperfect” may derive from Latin “imperfectus” meaning “unfinished”. It has long been regarded as a phenomenon unique to certain Indo-European languages such as Romance, Latin, Greek, Indo-Iranian and Balto-Slavonic languages [1]. However, Imperfect is semantically universal in human languages, whose meaning can be analyzed as the combination of imperfective and past tense [2]. When it comes to its specific expressions, however, it can vary across regions and languages, which are worthy of further study.

Through the analysis of the tense-aspect encoding devices of 105 languages samples collected, we propose three encoding types of imperfect: separative marker, cumulative imperfect marker and past tense marker. Among them, languages with cumulative imperfect markers can be interpreted to have more grammaticalized and distinctive marking of Imperfect. Thus, it is assumable that the usage of Imperfect category is more systematic and fully-developed in those languages, which serve as valuable resources for exploring semantics of Imperfect category.

DOI: 10.54254/2753-7064/3/20220237

2. Literature review

2.1. Definition of Imperfect

As there are few previous studies defining the term of "imperfect" explicitly, in this paper, a clear and comprehensive definition of the term "imperfect" are inspected from tense and aspect system respectively.

Tense is a deictic system which allows a speaker to locate a situation relative to some reference point in time [3], represented by Reichenbach's theory [4]. The imperfect verb form is mainly used to indicate the past situations. There is a tendency that the Perfective category to be restricted to past time reference, and the past is restricted to imperfective [2]. The past tense constitutes an essential part of the definition and meanings of imperfect.

Basically, aspect refers to grammatical or viewpoint aspect, and in general sense also includes lexical aspect (i.e. aktionsart). The lexical aspect depicts a situation's internal temporal structure. Five types of situations are distinguished: states, activities, accomplishments achievements [5] and semelfactive [6], which are classified by three features: stative, telic and duration [6]. The grammatical aspect depicts "different ways of viewing the internal temporal constituency of a situation" [1], as shown in Figure 1. The imperfective aspect can be divided into habitual and continuous. Habitual "describes a situation which is characteristic of an extended period of time" [7], while continuous means an event is in progress in a time point or period. Imperfect verb form expresses imperfectivity. Comrie concluded that imperfect form was inflected from the present stem with past tense morpheme. So, the imperfect was characterized as "present in the past" or "imperfective plus past time meaning" [1]. Dhal proposed a tripartite TAM system including Imperfect, Aorist and Present, as shown in Figure 2. In this sense, imperfect can be seen as the combination of imperfective and past [2].

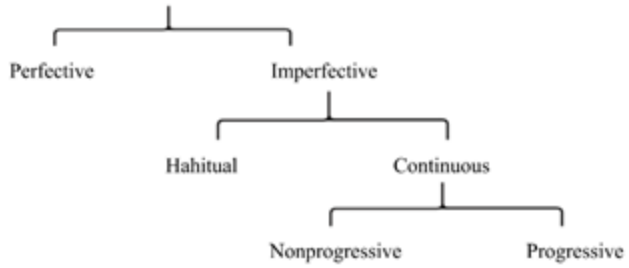

Figure1： Comrie's Classification of aspectual oppositions

Figure2： Dahl's Aspect System

Considering the foregoing discussion, in this study, Imperfect is defined as the imperfective aspect in the past tense, which exists as a cross-linguistic concept. The imperfect meaning can be expressed by periphrastic construction, single verb inflection or other encoding devices discussed later.

2.2. Research Questions

Three research questions are delivered throughout this paper:

DOI: 10.54254/2753-7064/3/20220237

(1) What are the encoding types of Imperfect?
(2) What are the possible reasons for the forming of the particular types?
(3) What are the universals and characteristics of these types?

In terms of the first question, this research paper has proposed three encoding types of Imperfect: separative marker, cumulative imperfect marker and past tense marker, on the basis of the analysis of the Imperfect encoding devices and characteristics of our language samples, combined with relevant theories of the encoding types of tense and aspect

With regard to the second question, explanations from several aspects both inside and outside the language system are offered, including the inflectional synthesis of the language, the prominence of tense and aspect, and external historical factors

As for the third question, the common ground and characteristics of these types are elaborated from the semantical perspective. It claims that there is a difference in the degree of development of Imperfect category between the types, among which languages with cumulative Imperfect marker have the most developed and relatively independent Imperfect category.

3. Research Method

To analyze the cross-linguistic encoding devices and semantics of imperfect, language samples are established. In this process, the genetic bias and the areal bias should be avoided [8], to ensure that each language family and area be represented equally and adequately by the sample language chosen. Languages are selected proportionally according to their geographical distributions. To ensure the chosen languages are geographically representative, division of language in macro-area from The World Atlas of Language Structures (WALS) is adopted. 105 languages are chosen as sample languages, as shown in Table 1.

Table1: The geographic proportional selection of languages

Area	Total Languages	Chosen Languages
Africa	604	24
Australia	177	7
Eurasia	654	26
North America	394	16
Papunesia	557	22
South America	255	10

4. Results

4.1. Overview

The encoding devices of imperfect vary in the world languages. In an attempt to divide the types of imperfect encoding devices across languages, it is necessary to start with the cross-linguistic encoding devices and encoding types of tense and aspect. In this section, tense and aspect encoding devices in our language samples are summarized, based on which the types of imperfect encoding devices are generated.

Dryer proposed three types of tense and aspect morphemes: affixes, particles, and auxiliaries [9]. Affixes, including prefixes and suffixes, can be taken as a form of inflection on the word stem. Auxiliaries and verb particles are two main categories of function words related to verbs, the former with inflected forms while the latter without. In addition, tense-aspect information can also be expressed by anti-passive voice, which is categorized as affixes in this paper. Tones can also be used

DOI: 10.54254/2753-7064/3/20220237

as tense-aspect encoding devices. However, according to WALS, only 13 out of 1131 recorded world languages use tones as tense-aspect marker. Considering that, it is not included in this paper.

In our sample languages, 57 out of 105 languages tend to use the same encoding devices to indicate both tense and aspect, while nearly the other half do not. Given that, the encoding devices for tense and aspect are recorded respectively.

Table 2 shows the numbers of languages using certain devices to encode tense and/or aspect. Inflection is the major one, with nearly 85% of the language samples having inflection as at least one of its tense and aspect encoding devices. Particles and auxiliary verbs can be used as a complement to infection, or as an independent and complete system to encode tense and/or aspect. If a language is not that rich in morphology, it can turn to content words to express tense and aspect information.

Table2: Tense and Aspect Encoding Devices

	Inflection	Particle	Auxiliary	Content Word
Tense	78	17	14	18
Aspect	66	14	17	9
Either	89	21	22	18

Tense-aspect encoding systems can be distinguished from the perspective of exponence [10]. Monoexponential languages have separative formatives, that is with a single grammatical property each; polyexponential languages have cumulative formatives, each with several grammatical properties, such as gender, number, case, tense and aspect. In terms of tense and aspect, correspondingly, languages can have separative and cumulative formatives: the former refer to two separable formatives expressing tense and aspect respectively; the latter refer to one single formative containing information of both categories simultaneously. This is an important classification criterion for encoding types of Imperfect below.

4.2. Encoding Devices of Imperfect

Among the 105 language samples, common encoding devices for Imperfect include inflection, auxiliaries, particles and content words.

Inflection, including affixation and non-affixation, is found as the most common encoding device of imperfect. For example, in Lango, a language with 6 to 7 morphemes per word, the imperfect meaning is marked by inflection. As shown in (3), verbs are inflected to indicate both progressive and perfective aspect.

(1) Lango [11]
òŋwòŋò lócə àcɛm
find.3SG.PERF man eat.3SG.PROG
The man was eating.

An auxiliary verb adds grammatical meaning to the main verb. Both auxiliary verbs and particles only have grammatical functions, but no lexical function. Particles differ from auxiliary in that they do not have inflection forms. For example, in (2) a, Chamorro uses the particle “Ginin” to express the imperfect meaning. By contrast, in modern Hebrew (2) b, the past tense is encoded with the help of the auxiliary verb “haye”.

(2) a. Chamorro [12]
Ginin manhohokka
IMPF AGR.ANTIP.collect.PROG
yu’ bronsi anai didikiki’ yu’.
I bronze when AGR.small.PROG I
I used to collect bronze when I was small.

DOI: 10.54254/2753-7064/3/20220237

b. Hebrew (Modern) [13]
ha-yeladot hayu roqdot be-ma'agal
The-girl. BE.PST. dance.PTCP. in-circle
The girls used to dance in a circle.

Apart from the encoding devices with relatively high degree of grammaticalization mentioned above, content words can also express the imperfect information. In many languages which adopt other more grammaticalized encoding devices, content words are often used to provide supplementary tense and aspect information. In those languages, the imperfect information can still be expressed completely, even if the content words are removed. However, content words can also be used as the primary (but not obligatory) encoding device in some languages, such as Indonesian and Chinese, largely depending on the language's index of synthesis (the average number of morphemes per word). The imperfect meaning, which depends partly on inference or implication, cannot be expressed clearly without content words in those languages. For example, in Indonesian (3) a, if not further explained, the verb form alone can indicate either the present or the past tense, hardly expressing the clear imperfect information. By contrast, in (3) b, both the temporal noun "kemarin" and the adverb modifier "tadi" are used as the content words to identify the absolute or relative time to make the imperfect meaning clear.

(3) Indonesian [14]
a. Mereka sedang makan.
3PL keep.PTCP.PROG eat
They are/were eating.
b. Mereka sedang makan
3PL keep.PTCP.PROG eat
kemarin/tadi.
yesterday/recently.
They were eating yesterday/a short time ago.

4.3. Encoding Types of Imperfect

Based on the analysis of the Imperfect encoding devices and characteristics of our language samples, three encoding types of Imperfect are proposed and explained: separative marker (the past tense and the imperfective aspect marked respectively), cumulative imperfect marker (specific marker) and past tense marker (past tense with imperfect meaning).

As have been mentioned above, separative formatives for tense and aspect contain two separable formatives expressing tense and aspect respectively. Since Imperfect contains the meaning of Past tense and Imperfective aspect, about 81% of the sample languages tend to use two morphemes to express the tense and aspect information respectively and together they form the complete meaning of Imperfect. For example, in Hindi (4) a, the auxiliary verb "the" denoting the past tense and the imperfective form of the verb "ga:rahe" together make the marker of imperfect. Cumulative imperfect marker is a single inseparable morpheme with complete Imperfect meaning. For example, in Spanish (4) b, the verb inflection form "hacía" itself contains both the meaning of Past tense and Imperfective aspect. The last encoding type is relatively rare and not representative enough in human languages, only found in Hebrew and Persian among our language samples, which use past tense marker to imply Imperfect meaning.

(4) a. Hindi [15]
bacce sku:l ja:tee hue
Children school go-while
ga:rahe the.
sing-PROG BE.PST.PL

DOI: 10.54254/2753-7064/3/20220237

```
The children were singing songs while going to
school.
b. Spanish
Iba                  a     la       escuela
go.IMPF.1SG   to    ART    school
cuando    llegaste.
when       come.PAST.2SG
I was going to school when you came.
```

The first two of the encoding types of Imperfect (separative marker and cumulative imperfect marker) are largely attributable to the inflectional synthesis of the language. First of all, if there are not enough inflection categories in a language (more of an isolated language), it is more likely for its tense and aspect to be encoded in different ways, one by inflection and the other by an auxiliary verb or function word for example, as in English and Hindi. Then, if there are relatively more inflection categories in a language (more of a synthetic language), it is more likely to have cumulative formatives to express the Imperfect information, namely using a single, specialized affix or non-affix inflection form to encode Imperfect. Finally, if a language is rich in its inflection categories (more of an agglutinative language), it is probable to add two affixes denoting the tense and aspect respectively to a word stem to express the complete meaning of imperfect. As two of the three cases above belong to the type of separative marker, that may explain why separative marker is found as the dominant encoding type of Imperfect among the language samples.

The last type can be explained from the perspective of the prominence theory by Bhat. The prominence of a language form can be expressed through the degree of grammaticalization, obligatoriness, paradigmatization and pervasiveness [16]. According to our language samples, zero-marking, namely the past tense marker with imperfect meaning, are most common in Semitic languages which are traditionally regarded as tense-prominent languages. Compared with aspects, these languages attach the relative prominence to tense by grammaticalizing the category to a greater degree and by making it more obligatory and more systematic. In those languages, the verbs themselves can contain the Imperfective information. As a result, their past forms automatically express the meaning of both past tense and imperfective aspects, namely the meaning of Imperfect.

4.4. Languages with Cumulative Imperfect Marker

After the division and analysis of the encoding types, it is plausible to claim that the Imperfect category is more systematic and fully-developed in languages with cumulative imperfect marker considering its more grammaticalized and distinctive marking of Imperfect. (They do not necessarily use a cumulative formative to encode other combinations of tense and aspect.) Thus, this particular type, which has traditionally been regarded as the language that Imperfect is particular to, requires additional attention and discussion to reach a better understanding of Imperfect as an independent category.

Among the 105 language samples established, 18 languages have cumulative Imperfect marker, which accounts for about 17%. 13 out of these 18 languages are distributed in Eurasia area, to be specific, in the west of the Middle East. Among other areas, South America and Papunesia have three and two languages of this encoding type respectively.

Languages with cumulative Imperfect marker generally show two tendencies: Genetically, they are concentrated in Indo-European languages, especially Romance languages; geographically, four of them are concentrated in the Caucasus, with others scattered around the world.

Modern Romance languages are developed from Vulgar Latin, in which the Imperfect category played an important role, in the expansion of Roman Empire. In Latin, the “imperfectus” is one of the past tenses, with a common verb ending “-bam”. Modern Romance languages are more or less

DOI: 10.54254/2753-7064/3/20220237

similar to Latin in the Imperfect category and its inflection rules. The Imperfect category in modern Greek also has its traces in ancient Greek. Thus, it can be deduced that those languages with cumulative Imperfect marker can inherit the Imperfect category from some influential ancient languages such as ancient Latin and Greek.

Languages around Caucasus belong to three different and unrelated language families: Northeast Caucasian languages (Nakh-Daghestanian), Northwest Caucasian and South Caucasian languages (Kartvelian). Languages with cumulative Imperfect markers are gathered here, which may be more of a geographic pattern than a genetic pattern. Even though the speculation of a common Proto-language is hardly enough to offer a definitive answer, the Caucasus is still expected to find more languages with cumulative Imperfect markers.

In our language samples, two-thirds of languages with cumulative Imperfect markers are found in Eurasia area, and most of them are found in Indo-European languages, roughly consistent with Comrie's findings on Indo-Iranian, Greek, Latin and Balto-Slavic languages. In seven major groups of modern Indo-European languages, our language samples from Italic, Hellenic, Baltic, Slavic and Indo-Iranian languages have all been confirmed with cumulative Imperfect marker, but none in Celtic and Germanic languages. Outside the Eurasia area, three languages from South America and two languages from Papunesia have been found with cumulative Imperfect markers, but there are no distinct genetic or geographic relations between them.

5. Discussion

5.1. Basic Semantic Features

There is a common ground from which the encoding types are divided in this study, that is Imperfect as a semantically universal concept across languages, instead of a regional term. Therefore, it has basic semantic features shared by most language samples, regardless of the different encoding devices. With reference to the definition of Imperfect mentioned before, Imperfect is typically used to express past habitual and progressive events. Habitual describes the characteristic situation in the past period of time, often including Iterative action, and progressive describes the past on-going event.

5.2. Unique Semantic Features

As mentioned before, languages with cumulative Imperfect markers can be regarded as a special type with the relatively independent Imperfect category. One of the reasons or manifestations is that apart from the basic semantic features, they have diversified use of Imperfect. Here are some of their unique semantic features found in our language samples.

5.2.1. Past Static Description

Imperfect is compared with other aspects to explain one of its unique semantic features to express past static situation. In Indo-European languages, Perfective category tends to be bounded to the past time reference [2], that is a completed action typically taking place in the past time. By contrast, Imperfective often goes together with simple present tense and is formed by the present stem. If a language does not have compulsive verb inflection distinguishing imperfective and perfective, the present state in that language is typically indicated by the present verb form instead of the imperfective form. It is assumable that a state is typically indicated by a simple tense inflection instead of aspect inflection. However, in our language samples, some languages with cumulative Imperfect marker choose the imperfect form rather than the simple past form to describe the state in the past.

(5) Ingush [17]

a. Aaz uqaza jeaqqaa jerriga

DOI: 10.54254/2753-7064/3/20220237

1s.ERG here J.spend.PPL J.all
xa sixa dwa-jaxar.
time fast DX-J.go.WP
The time I spent here passed quickly.
b. Qy yz maara hama xaacar cynna
else this except thing know:NEG.IMPF 3s.DAT
He knew nothing but this. This is all he knew.

For example, in Ingush, the dynamic verb "go" in the past in (5) a is marked by witnessed past tense (simple past tense), while the static verb "know" in (5) b is marked by imperfect form instead of simple past form.

5.2.2. Background

The opposition between Perfective and Imperfective aspects can distinguish foreground and background in a sentence. The perfective aspect corresponds to the foreground, often used to indicate the main event or situation, while the imperfective aspect corresponds to the background, providing secondary background information [18]. The imperfect form can be used to express the background event in the past.

(6) Italian [19]
a. Marco ha telefonato quando
Marco have-3SG telephone-PAST when
eravate al cinema.
BE-2SG-IMPF at.the cinema.
Marco phoned when you were at the cinema.
Portuguese [20]
b. Eu dormia quando tu chegaste.
1SG sleey-IMPF when 2SG get-PAST
I was sleeping when you arrived.

This semantic feature of Imperfect is found in most Romance languages with cumulative Imperfect marker. In the Italian example 6 (a), "you were at cinema" is a continuous situation, which can be seen as background information, and "Marco telephoned" is an incident happening against the background, which can be as the foreground. The background verb "be" takes the imperfect form, while the foreground verb "telephone" takes the simple past form. The same applies to Portuguese. In the Portuguese example, "I was sleeping" is the background situation expressed by the imperfect form, while "you arrived" is the foreground incident expressed by the simple past form.

5.2.3. Counterfactuality

In our language samples, the semantic feature of expressing counterfactuality is mainly found in Romance languages, Indo-Iranian languages, and languages in the Caucasus. These languages with cumulative Imperfect marker use the imperfect form to express counterfactuality.

(7) Abkhaz [21]
a. arə´j ø -z-də´r-wa-za+r
this it-1-know-PRES:DYN-COND
ja-q´a-s-c´a-wá+n
it-PREV-1-ROOT-IMPRF
(Even) if I knew this, I would have done it.
b. arə´j ø -wa-də´r-wa-z+t+g'ə
this it-you-know-even if
jə-wə-zə´-q'a+c'a-wa+mə+z+t'

DOI: 10.54254/2753-7064/3/20220237

```
    it-you-POT-do-NEG.IMPRF
    Even if you had known this, you could not have
done it.
Ingush [17]
c. Sie          cynna        bwarjga+guljga
   1SG.RFL   3SG.DAT   eye+see.SBJ
loura           suona
want.IMPF   1SG.DAT
    I'd like him to see me (myself).
```

For example, in Abkhaz 7 (a) (b), the imperfect form is used in the conditional clause to indicate the counterfactual situation in the past. In Ingush example 7 (c), the imperfect form of "want" expresses the wish which is not likely to be true in the future.

6. Conclusion

With the establishment of 105 language samples across world, we have conducted a detailed analysis of Imperfect category from a typological perspective. The imperfect meaning is encoded variously across languages, and thus can be divided into three different encoding types: separative marker, cumulative imperfect marker and past tense marker. Based on the typological classification of the encoding types, the semantic features of Imperfect are investigated discriminately. Corresponding to the definition of Imperfect in this study as the combination of imperfective aspect and past tense, its universal semantic feature involves expressing the habitual event or on-going action in the past. However, its semantic features are far more than that. In languages with cumulative Imperfect marker, which are regarded to have the most developed and systematic Imperfect category, the imperfect verb forms are found to be able to indicate the past state, background, and counterfactuality. Such typological study on Imperfect demonstrates how a synthesis of tense and aspect can make a difference on its basic meanings and uses in certain languages, obliging linguists to adapt their protocols and methodologies to fully exploit the rich potential of Imperfect as a special and independent category.

Acknowledgments

Both authors listed have contributed equally to this work, and should be considered co-first authors.

References

[1] *Comrie, B. (1976). Aspect. Cambridge: Cambridge University Press.*
[2] *Dahl, Ö. (1985). Tense and Aspect Systems. Oxford: Blackwell.*
[3] *Saeed, J. (2009). Semantics (3rd ed.). Oxford: Blackwell.*
[4] *Reichenbach, H. (1947). Elements of Symbolic Logic. New York: Macmillan.*
[5] *Vendler, Z. (1967). Verbs and Times. In Linguistics in Philosophy, New York: Cornell University Press, 97-121.*
[6] *Smith, C. S. (1991). The Parameter of Aspect. Dordrecht: Kluwer.*
[7] *Comrie, B. (1985). Tense. Cambridge: Cambridge University Press.*
[8] *Bell, A. (1978). Language Samples. In J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (eds), Universals of Human Language. Cambridge, MA: MIT Press, 123-56.*
[9] *Dryer, M. (1992). The Greenbergian Word Order Correlations. In Language, 68, 81-138. https://www.jstor.org/stable/416370*
[10] *Bickel, B. & Nichols, J. (2013). Exponence of Selected Inflectional Formatives. In Dryer, M. & Haspelmath, M. (eds.), The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/21*
[11] *Noonan, M. (1992). A Grammar of Lango. Berlin & New York: Mouton de Gruyter*
[12] *Chung, S. (2020). Chamorro Grammar. Santa Cruz: University of California. http://dx.doi.org/10.48330/E2159R*

DOI: 10.54254/2753-7064/3/20220237

[13] Boneh, N. (2013). Tense: Modern Hebrew. In Khan, G. (eds), Encyclopedia of Hebrew Language and Linguistics. Leiden: Brill. http://dx.doi.org/10.1163/2212-4241_ehll_EHLL_COM_00000135
[14] Sneddon, J. N. (1996). Indonesian: A Comprehensive Grammar. London/New York: Routledge.
[15] Koul, O. (2008). Modern Hindi Grammar. Springfield: Dunwoody Press.
[16] Bhat, D.N.S. (1999). The Prominence of Tense, Aspect and Mood. Amsterdam/Philadelphia: John Benjamins.
[17] Nichols, J. (2011). Ingush Grammar. Berkeley & Los Angeles: University of California Press.
[18] Hopper, P. J. (1979). Aspect and foregrounding in discourse. In T. Givón (ed.), Discourse and Syntax, Vol. 12. New York: Academic Press, 213-241.
[19] Peyronel, S. & Higgins, L. (2006). Basic Italian: A Grammar and Workbook. London & New York: Routledge.
[20] Hutchinson, A. & Lloyd, J. (1996). Portuguese: An Essential Grammar. London & New York: Routledge.
[21] Chirikba, A. (2003). Abkhaz Grammar. München: Lincom GmbH.