Automatic Detection of Conceptual Time Metaphors

Seong-Eun Cho

Masteroppgave i datalingvistikk og språkteknologi
Institutt for lingvistiske, litterære og estetiske studium
Universitetet i Bergen
2013

UNIVERSITETET I BERGEN

UNIVERSITETET I BERGEN

Institutt for lingvistiske, litterære og estetiske studium

DASP350
Masteroppgave i datalingvistikk og språkteknologi
Vårsemester 2013

Automatic Detection of Conceptual Time Metaphors

Seong-Eun Cho

Abstract

This thesis presents a system for the automatic detection of conceptual metaphors of time. Metaphors representing time as space, such as "Time Orientation", "Moving Time" and "Moving Ego", are common, but not equally universal among different languages. A first attempt to recognize these metaphors is made with the development of the *CTM Detector* ("Conceptual Time Metaphor Detector"). This program analyzes English text and detects specific grammatical combinations of time expressions and relevant verbs with semantics mainly related to movement and sight. The program has been trained and tested on a corpus of political speeches by president Barack Obama. The evaluation shows that this program manages to retrieve linguistic realizations of the metaphors with an overall precision and recall of 89.17% and 88.50% respectively. The detection and processing of conceptual metaphors could improve natural language applications such as machine translation, since current statistical machine translation systems often produce unsatisfactory outputs when certain types of conceptual metaphors, e.g. "Moving Ego", occur in the source language but are avoided in the target language.

Sammendrag

Denne oppgaven presenterer et system for automatisk gjennkjenning av konseptuelle metaforer mht. tid. Metaforer som fremstiller tid som rom, f.eks. "Tid som orientering i rom", "Tid som beveger seg" og "Ego som beveger seg i tid" finnes ofte, men de er ikke like universelle blant ulike språk. Et første forsøk til gjenkjenning av disse metaforene består av utviklingen av systemet *CTM Detector* ("Konseptuell tidsmetafordetektor"). Dette programmet analyserer engelsk tekst og finner spesifikke grammatiske kombinasjoner av tidsuttrykk og relevante verb med et semantisk innhold hovedsakelig relatert til bevegelse og syn. Programmet ble trent og testet på et korpus av politiske taler av president Barack Obama. Evalueringene viser at programmet kan gjenkjenne korrekte lingvistiske realiseringer av metaforene med en total presisjon og fullstendighet på henholdsvis 89.17% og 88.50%. Gjenkjenning og prosessering av konseptuelle metaforer kan forbedre språkteknologiske anvendelser som maskinoversettelse, fordi dagens statistiske oversettelsessystemer ofte gir lite tilfredsstillende resultater når visse typer av konseptuelle metaforer, f.eks. "Ego som beveger seg i tid", forekommer i kildespråket, men ikke er gangbare i målspråket.

Preface

"How do you translate the sentence 'We are approaching Christmas' in German, Mr. Cho?"
"Wir nähern uns Weihnachten."
"And how do you translate this sentence in Korean?"
"I'm confused ..."

My former lecturer, Dr. David West, must have been very puzzled about my answer. I was confused because my feelings told me that translating this example literally, results in having a non-acceptable Korean sentence. This was in fact the key moment when I got interested in conceptual metaphors and how time is expressed in different languages. I therefore want to thank him for the inspiration and introducing me to theories of Cognitive Linguistics. I also want to thank Dr. Martin Hoelter for extending my knowledge on metaphor theories, and Jan Strunk for the great courses on programming in Perl. Ever since I took these courses on metaphors and programming it was my wish to test the possibility of automatically detecting conceptual metaphors and to discover solutions for improving machine translation systems. My most profound thanks are dedicated to Prof. Koenraad de Smedt who has guided me through the Master's program as my mentor, supervisor and lecturer. Without his willingness to accepts this great but also risky challenge to enter an almost untouched research area my wish would have not come to fulfillment. I am grateful for his encouraging feedback and constructive criticisms that raised the quality of the paper immensely. I want to thank all my friends and fellow students for making the past two years in Bergen so memorable. Special thanks go to my friend Magnus Bakken for lending me his iMac to finish my thesis, after an entire glass of water was spilled right on the keyboard of my five-month-old Macbook Air. Last but not least, I want to thank the most important people in my life. I want to thank my parents and my sister for their great support, prayers and love.

Contents

Chapter 1

Introduction

The concept of time has been thoroughly studied in various disciplines over a long period of time. One of the first scholars who tackled this topic was the Latin philosopher St. Augustine who concluded with the famous quote: "*Quid ergo est tempus? Si nemo ex me quaret, scio. Si quaerenti explicare velim, nescio.*"[1] This quote illustrates perfectly the problems ancient scholars had with conceptualizing time. By now, many studies from the fields of philosophy, psychology and linguistics have elaborated on this matter and taken different approaches to analyze specific aspects of this notion. One of the more recent linguistic frameworks that dealt with time is called the *Conceptual Metaphor Theory*. Lakoff and Johnson first introduced the concept of *Conceptual Metaphors* in "Metaphors We Live By" (Lakoff and Johnson, 1980) and they argue that abstract concepts receive their knowledge structures from other sensory-motor domains. Lakoff, Johnson and other scholars (Jackendoff, 1983; Langacker, 1987) agree that "space and time seem to show a peculiar relatedness that is perhaps not evident to a naive philosophical observer: Human languages again and again express temporal and spatial notions in a similar way" (Haspelmath, 1997). In order to describe temporal experience human beings use terminologies and concepts from the domain of space: "We are approaching the end of the game". Findings like these have been discussed from a theoretical perspective within the fields of *Cognitive Linguistics*. One of the major tasks of cognitive linguists was to verify the existence of a conceptual metaphor by e.g. manually searching for their surface realizations in corpora: "[I]t is necessary to list potential linguistic realizations and then trawl concordance lines to see if they occur. […] Once retrieved, a concordance will show the researcher the linguistic contexts in which a lexical item is used, but this information then has to be processed manually." (Deignan, 2005). Other researchers even recommend to either manually go through an entire, smaller corpus (Cameron and Deignan, 2003) or to analyze a subpart of a larger corpus (Charteris-Black, 2004). All of these approaches to discovering conceptual metaphors are time-consuming and require patience, nerves of steel and lots of coffee.

So far, no attempt has been made to automatically retrieve conceptual metaphors in the domain of natural language processing. One of the questions that will be tackled in this paper is whether it is really true that "the computer cannot work from a list of conceptual metaphors to identify their linguistic realizations" (Deignan, 2005). Doing research on all existing conceptual metaphors would have been too ambitious and unachievable within the given time. For this reason, the research scope is narrowed down to a list of three conceptual metaphors that are based on time: Time Orientation, Moving Time and Moving Ego. The thesis will present a program called "CTM Detector" (abbr. of "Conceptual Time Metaphor Detector") that is developed to identify and return linguistic realizations of these time metaphors. So far, the Conceptual Metaphor Theory has only been regarded as a theoretical framework that has no relevance for processing natural language data. The major motivation behind this project is the strong conviction that the automatic detection and processing of conceptual metaphors can improve natural language applications such as machine translation.

The structure of this paper is as follows: A brief account of the Conceptual Metaphor Theory introducing concepts and technical terms that are used throughout the paper is given in Chapter 2. Chapter 3 describes the goals that were set prior to developing the CTM Detector and it outlines the challenging aspects of approaching this research project. Afterwards, the components of the CTM Detector are introduced in Chapters 4 and 5. Chapter 4 presents the preprocessing modules of the CTM Detector that tags and parses sentences.

[1] Translation: "What, then, is time? If no one asks me, I know. But if I want to explain it to someone, I do not know."

The component that extracts instances of conceptual time metaphors is described in Chapter 5. First, it gives a description of the preliminary version of the extractor (beta version) containing four basic extraction rules. These rules were thought to capture most of the conceptual time metaphors. The evaluation of the beta version showed however that modifications and extensions were required. This chapter is concluded with an account of the final version of the CTM Detector that is extended by six additional rules. Chapter 6 presents statistics about the performance of the CTM Detector, and Chapter 7 discusses the importance of identifying and processing conceptual metaphors in the fields of natural language processing.

The CTM Detector and related files described in this paper will be made available through BORA and the CLARINO infrastructure.

Chapter 2

Conceptual Metaphor Theory

The common-sense notion of the term "metaphor" is connected to the rhetoric device of expressing creative, figurative language. The etymological root of "metaphor" is found in the Greek word "metaphora" which can be literally translated to "transfer". The greek philosopher Aristotle is known to be the first person who wrote a scholarly account on this topic. In *"Poetics"*, he argued that "[m]etaphor is the application of a strange term either transferred from the genus and applied to the species or from the species and applied to the genus, or from one species to another or else by analogy" (Aristotle et al., 1995). From Aristotle's perspective, sentence (2.1) is metaphorical and is not to be understood literally, as a human being cannot be an animal. Instead, certain salient defining characteristics of the concept "bear" are "transferred" to the entity "Mike Tyson" and the listener or reader understands that Mike Tyson's strength is being emphasized as in paraphrase (2.2).

(2.1) MET: Mike Tyson is a bear.

(2.2) PAR: Mike Tyson is strong.

In Searle's (1993) pragmatic account on metaphor, the "speaker's utterance meaning" can be derived from the literal, "sentence meaning" by testing six principles. Processing metaphorical expressions takes longer than non-metaphorical ones due to the necessity of additional cognitive processes.

This thesis is based on the framework called "Conceptual Metaphor Theory" and was introduced by the cognitive linguists Lakoff and Johnson (1980). One of the core statements of this study is that "[o]ur ordinary conceptual system [...] is fundamentally metaphorical in nature" (Lakoff and Johnson, 1980). The groundbreaking point is that the locus of metaphor is shifted from language to thought. Metaphors are regarded as a conceptual rather than a linguistic phenomenon and metaphorical utterances are not individual cases, which have to be analyzed one by one. The common-sense notion of metaphor as a rhetoric device is opposed and Searle's statement that it took longer to process metaphorical utterances is rejected. In this framework, metaphor is viewed as a structure-giving phenomenon firmly established between pairs of conceptual domains, or knowledge representations. Very often one domain is understood in terms of another. In everyday life, we utter sentences such as (2.3) which can be paraphrased as (2.4).

(2.3) I **see** what you mean.

(2.4) I **understand** what you mean.

Thus, we choose the lexical items, knowledge and inference structures of one domain (Seeing) in order to express those of another (Knowing). These systematic correspondences between two domains are called 'cross-domain mappings'. In the "KNOWING IS SEEING" metaphor, KNOWING is the target and SEEING the source domain. A characteristic of these conceptual metaphors is that source domains are predominantly derived from sensory experience whereas the target domains contain abstract concepts. "[M]any of the most basic concepts in our conceptual systems are also normally comprehended via metaphor – concepts like time, quantity, state, change, action, cause, purpose, means, modality and even the concept of a category" (Lakoff, 1993). These abstract concepts are difficult to describe in their own terms as they lack physical characteristics.

Various scholars have observed that the concept of time is mostly described with non-temporal words. The quotes by Haspelmath (1997), Clark (1973), Langacker (1987) and Jackendoff (1983) support this statement:

- "[S]pace and time seem to show a peculiar relatedness that is perhaps not evident to a naive philosophical observer: Human languages again and again express temporal and spatial notions in a similar way" (Haspelmath, 1997)
- "For a long time, linguists have noted that the spatial and temporal terms in English and other related languages overlap considerably" (Clark, 1973)
- "[W]e often conceive and speak of time in spatial terms" (Langacker, 1987)
- "[P]repositions of time are on the whole identical to spatial expressions and that temporal PPs are attached to sentences in the same way as PPs of location." (Jackendoff, 1983)

In conclusion, the experience of time seems to be dependent on spatial awareness. The following sections will describe how the concept of time is understood within the Conceptual Metaphor Theory.

2.1 Properties of Time

Lakoff and Johnson (1999) argue that the human body does not have a sensory-perceptual system to measure and perceive time "in itself". In order to measure the duration of certain events human beings are dependent on man-made instruments that display continuously iterated events, also referred to as "time-defining events". The most used instrument to measure time is the clock. The second hand of the clock moves along bit by bit at one-second intervals and users can e.g. measure the duration of an event by counting the number of these second events. Time's dependency on events also has the consequence that the properties of events are projected to the following properties of time (Lakoff and Johnson, 1999):

1. Time is directional and irreversible because events are directional and irreversible; events cannot "un-happen."
2. Time is continuous because we experience events as continuous.
3. Time is segmentable because periodic events have beginnings and ends.
4. Time can be measured because iterations of events can be counted.

So the brain gathers visual and other sensory information, isolates single events by determining a starting and an ending point, and measures its length. In order to segment a continous string of happenings into single events another factor needs to be taken into consideration. The transition from one event to the next is determined by changes occuring through modifications of situations. In most cases changes are perceived visually through *motion*. The time it takes to get from home to university or the time it takes for a friend to come to one's own apartment is measured by the shift from being stationary to moving, and back to being stationary in another location. According to Lakoff and Johnson (1999) time is tightly interwoven with motion as the experience of time is mostly grounded in the experience of motion events.

2.2 Conceptual Time Metaphors

Lakoff and Johnson (1999) introduce three different types of conceptual time metaphors which are based on the experience of event, motion and space.

1. Time Orientation Metaphor
2. Moving Time Metaphor
3. Moving Ego Metaphor

The first metaphor forms the foundation for the second and third metaphor. The latter two metaphors are extended versions of the Time Orientation metaphor as they integrate the concept of motion. These three time metaphors play a central role in this thesis as their linguistic realizations will be automatically extracted by the program introduced in the chapters *CTM Detector: Preprocessing* (Chapter 4) and *CTM Detector: Extraction* (Chapter 5).

2.2.1 Time Orientation Metaphor

The components of the Time Orientation metaphor are the speaker, referred to as the "ego", and a firmly determined sequence of time instances. These two components are organized as illustrated in Figure 2.1.

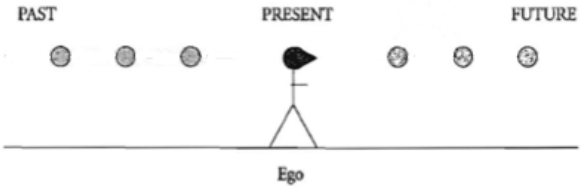


Figure 2.1: Time Orientation Metaphor (Modified Version of Fig. 2.2)

The ego forms the center of this metaphor and is depicted as a stick-figure who can look in a certain direction. He faces the future, the location of the ego represents the present and the space behind him forms the past. Time events are represented as round objects that are arranged on an ordered line (vector). Those objects that are in front of the ego represent events lying in the future, and the closer they lie to the ego the closer they are to the present. Objects located behind the ego lie in the past, and the further away they lie the more time has elapsed. The entire model is stationary and linguistic realizations of this metaphor do not refer to any movement (2.5 – 2.7).

(2.5) "I look forward to a great future for America" – John F. Kennedy

(2.6) "Look back, and smile on perils past" – Walter Scott

(2.7) "What lies behind us and what lies ahead of us are tiny matters compared to what lives within us." – Henry David Thoreau

These examples illustrate that the ego can look in both directions. The ego can either face the future (2.5) or reflect about occurrences that happened in past by turning around (2.6). Quote (2.7) is basically a description of the structure of the Time Orientation metaphor. It describes the default setting of past events lying behind and future events lying in front of the ego.

2.2.2 Moving Time Metaphor

The Moving Time metaphor is based on the Time Orientation metaphor as both of them consist of the same components and are organized in the same manner. In contrast to the Time Orientation metaphor, the Moving Time metaphor contains the concept of motion. The time objects within this model approach the ego from the future, run past the ego and disappear behind his back. Motion is represented as arrows in Figure 2.2.

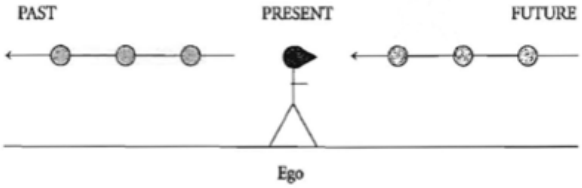


Figure 2.2: Moving Time Metaphor (Evans, 2004)

This modification gives rise to linguistic realizations in which temporal concepts are represented as objects that get closer to the ego or move away from him behind his back (2.8 – 2.10).

(2.8) "Nothing is more powerful than an idea whose time has come." – Victor Hugo

(2.9) "The present is a point just passed." – David Russell

(2.10) "The time is gone, the song is over, though I'd something more to say." – Pink Floyd

What all these examples have in common is that the temporal expressions are perceived as independent objects or "points" in time. The point in time *comes* closer to the ego (2.8), *passes* him (2.9) and is *gone* (2.10), behind his back. This time experience is closely related to the spatial experience of waiting for somebody. The person who is stationary and waiting for the other person forms the ego. The other participant approaches him from the far distance and as time passes the distance shrinks until to the point where they meet. After the meeting the person disappears behind the ego and distances himself more as time goes by.

The alternative to conceptualizing time as single objects or "points" in time, is to perceive time as a mass. This modified version is referred to as the "Time-Substance Variation" and it denotes a neverending mass continously running past the ego. Time is commonly conceptualized as a liquid mass flowing past the ego like a river. This statement is supported by quotes from the Roman emperor and philosophist Marcus Aurelius (2.11) and the Chinese philosopher Confucius (2.12).

(2.11) "Time is a sort of river of passing events, and strong is its current" – Marcus Aurelius

(2.12) "Time flows away like the water in the river" – Confucius

The first quote in example (2.11) illustrates clearly that it is the multiplicity of single, "passing events" that actually makes time appear like a mass. The overall picture of event sequences blurs and appears as a mass as many events overlap and similar events are merged together. Lakoff calls this phenomenon the "multiplicity-to-mass image-schema transformation" (Lakoff and Johnson, 1999). Examples of the Time-Substance variation are (2.13) and (2.14).

(2.13) "Time moves in one direction, memory in another." – William Gibson

(2.14) "Time goes on. So whatever you're going to do, do it. Do it now. Don't wait." – Robert De Niro

The quote in example (2.13) makes clear that time is directional and irreversible. Time's attribute of continuity in movement is emphasized in quote (2.14).

2.2.3 Moving Ego Metaphor

The Moving Ego (or Moving Observer) metaphor is also another version of the Time Orientation metaphor that is also extended with the concept of motion. In contrast to the Moving Time metaphor, the attribute of locomotion is not assigned to temporal events but to the ego. The ego is conceptualized as a non-stationary entity that moves on a landscape and walks past events that are "anchored" on a straight, horizontal path. The concept of PATH is an additional, essential component of the Moving Ego metaphor. Figure 2.3 visualizes the Moving Ego metaphor.

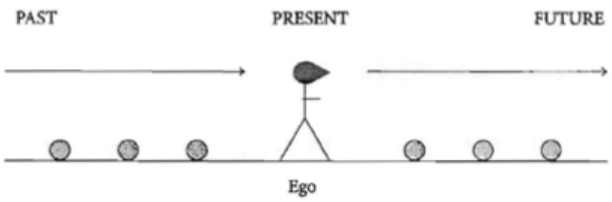


Figure 2.3: Moving Ego Metaphor (Evans, 2004)

The following examples given in this section reveal that a differentiation has to be made between two types of linguistic realizations of the Moving Ego metaphor. The quotes in the examples (2.15) and (2.16) show that the ego can get closer to a specific *point* in time.

(2.15) "I see a lot of signs every day that we're moving closer and closer to that tipping point." – Al Gore

(2.16) "Mr. President, we are rapidly approaching a moment of truth" – General "Buck" Turgidson in the movie Dr. Strangelove

The "tipping point" in example (2.15) and the "moment of truth" in (2.16) are moments in time. The experience of approaching these moments is derived from our spatial experience of getting closer to a stationary entity.

The next sentences (2.17 – 2.19) are different, as the metaphorical expressions are based on another type of spatial experience. Time units are not conceptualized as object-like entities but as a room or area that the ego can enter, go through and leave. The quote (2.17) reveals that the time noun "era" is described as a room that the ego can *enter*. Once entered the ego can walk through temporal space, expressed as *going through* "difficult times" (2.18). As soon as the end of the room is reached, the ego can *leave* temporal rooms such as "the 20th Century" (2.19).

(2.17) "…we could enter a new era of unlimited power that would do away with the need to dam our beautiful streams." – David R. Brower

(2.18) "When you are going through difficult times" – Paulo Coelho

(2.19) "Leaving the 20th Century" – Manic Street Preachers

Mentally or physically exhausting time periods are often represented in this manner (examples 2.20 – 2.22).

(2.20) "We ask for consideration and respect for our family as we go through this difficult time." (COCA)

(2.21) "What caused you to go through that dark period?" (COCA)

(2.22) "Between eight and twelve years old, kids go through an awkward stage." (COCA)

Taken together, the three conceptual time metaphors illustrate to what extent the experience of event, space and motion determine how time is conceptualized and modeled. The Time Orientation metaphor shows the structure of time by representing it in a systematic, spatial setting. Motion within this model can be assigned either to events (Moving Time metaphor) or the ego (Moving Ego metaphor).

2.3 Terminology: Time Senses

The previous section *Conceptual Time Metaphors* (Chapter 2.2) described the structure of the different types of conceptual time metaphors. The various examples that were given for each metaphor showed that temporal concepts can take different types of "shapes". Time can be conceptualized as a spatial room (as in examples 2.17 – 2.19), as a point in time (as in examples 2.15 or 2.16), or even as a liquid mass (as in example 2.12). Lakoff and Johnson (1999) set their major focus on the structure and attributes of conceptual time metaphors and paid less attention to these forms of time. Evans's *Structure of Time: Language, Meaning, and Temporal Cognition* deals with this matter and introduces eight different types of *time senses*, or lexical concepts of time. Evans explains that "time is associated with a large array or a semantic network of inter-related senses or lexical concepts – concepts stabilised in memory for the purposes of external representation via language." (Evans, 2004). This paper will not go into exhaustive detail of this study, but will focus on three major time senses that are most frequent: Duration, Moment and Matrix Sense.

Duration Sense

The Duration Sense refers to metaphorical expressions structuring time as the previously mentioned "spatial rooms". Evans (2004) introduces the concept of "time spans" being determined by two events that form the beginning and the end. "Duration Sense prompts for a lexical concept which constitutes an interval bounded by two 'boundary' events, i.e., the beginning and ending of the interval. I will define duration as the interval holding or extending between the two boundary (beginning and ending) events. I will term the beginning event the onset, and the ending event the offset." (Evans, 2004). This is illustrated in Figure 2.4.

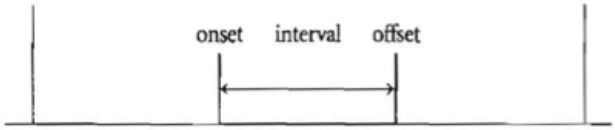


Figure 2.4: Duration Sense (Evans, 2004)

Moment Sense

The Moment Sense of time refers to a time expression that forms a specific "a discrete or punctual point or moment" (Evans, 2004) in time. The main difference between the Duration and the Moment Sense is that in the latter case durational readings are blocked entirely. The nouns "assignment deadline", "moment", "anniversary" and "dance competition" in examples (2.23 – 2.26) are perceived as discrete points in time.

(2.23) We are approaching the deadline quickly.

(2.24) The moment has come.

(2.25) We are getting closer to our 25th anniversary.

(2.26) The dance competition is drawing near.

In very special cases, discrete points in time can form the onset and the offset of time spans. A temporal "beginning" and "end" forms the onset and offset of time intervals respectively, as illustrated in (2.27) and (2.28).

(2.27) We have entered the beginning of an exciting season.

(2.28) He has reached the end of his life.

In the first example (2.27) the noun "beginning" forms the onset of the period "season", and in the second example (2.28) the "end" forms the the offset of a "life" span. According to Evans, temporal expressions that represent the onset or offset of a time span form an extra subcategory of the Moment Sense referred to as the "Event Sense". Evans justifies his decision to make this differentiation as "the Moment Sense references a temporal point (within a particular temporal event-sequence), [whereas] the Event Sense references an experiential point in an event-sequence." (Evans, 2004). For the sake of convenience both senses are merged together and will be simply referred to as the "Moment Sense" as both lexical concepts are perceived as discrete points in time.

Matrix Sense

In Lakoff and Johnson's "Time-Substance Variation" of the Moving Time metaphor, time is conceptualized as a mass which is created through the "multiplicity-to-mass image-schema transformation". The authors defined this form of time only very loosely as the "*the flow of time*" (Lakoff and Johnson, 1999). For this reason Evans' more technical term "Matrix Sense" will be used in later stages of the paper. The Matrix Sense of time refers to "an entity which is unbounded. In present terms we can say that in this sense time relates to an entity that it is not constrained by the interval holding between individual events, i.e., by an onset and offset [...]. As such, it indexes an entity which has an infinite elapse, and thus is conceived as subsuming all other events, the Matrix in terms of which experience is possible." (Evans, 2004).

2.4 Universality of Conceptual Time Metaphors

Lakoff and Johnson argue that "[t]he Moving Time and Moving Observer metaphors are not limited to English" and that "a preliminary survey suggests that these metaphors are common in the world's languages" (Lakoff and Johnson, 1999). The quotes and translations in (2.29 – 2.32) support their statement with examples from German.

(2.29) "**Das Beste sollte nie hinter uns, sondern immer vor uns liegen**." – Bertrand Russell
The best should never be behind us, but always ahead of us.

(2.30) "**Der zuverlässigste Weg, die Zukunft zu sehen, ist das Verstehen der Gegenwart**." – John Naisbitt
The most reliable way to see the future is by understanding the present.

(2.31) "**Wenn Weihnachten näher kommt, dann wird es heller in unserem Leben**." – Rainer Kaune
Our life brightens up when Christmas comes closer.

(2.32) "**Wir nähern uns dem Ende des Geldsystems, das 1971 eingeführt wurde**." – Bill Bonner
We are approaching the end of the monetary system that was introduced in 1971.

The first two examples (2.29) and (2.30) contain the primary dependencies of the Time Orientation Metaphor, the third quote (2.31) contains the Moving Time metaphor and the last sentence (2.32) reveals that Moving Ego constructions are acceptable in German.

However, the following inquiry on conceptual time metaphors in Korean will show that Lakoff and Johnson's assumption of the universality of the Moving Ego metaphor may be questioned. The Korean examples that will be presented throughout the rest of this section are all retrieved via the Korean search engine provider Naver[1] that accesses several, official English-Korean translation corpora[2]. The examples (2.33 – 2.35) show that expressions of the Time Orientation metaphor are found in Korean. Temporal concepts can lie ahead (2.33), behind the ego (2.34) and the ego can also face time events (2.35).

(2.33) ENG: … hard times which are **ahead**.
KOR: 앞에 놓인 어려운 시기
ap-e nohin eolyeoun sigi
front-LOC lie difficult time
"difficult time that lies **ahead**."

(*Source: The Telegraph*)

(2.34) ENG: The peak is now **behind us**.
KOR: 최악의 사태는 이제 우리 뒤에 남아있어요
choeag-ui satae-neun ije uli dwi-e nama-iss-eoyo
worst-POSS state-SUBJ.FOC now us behind-LOC remain-COP.PRS-POL (=Politeness)
"The worst state is **behind us** now."

(*Source: Neungyule Education*)

(2.35) ENG: We **face** many hard-to-make-a-decision moments in life.
KOR: 우리는 살면서 결정하기 어려운 많은 순간들에 직면합니다
uli-neun sal-myeonseo gyeoljeongha-gi eolyeoun manh-eun sungan-deul-e jigmyeon-habnida
we-SUB.FOC live-while decide-to difficult many moment-PL-ACC face-do.POL
"We **face** many hard-to-make-a-decision moments while living."

(*Source: TIMES CORE*)

Moving Time metaphors were frequently found in Korean, one of them being example (2.36). The English sentence was translated literally to Korean.

[1] `http://endic.naver.com/?isEngVer=Y`

[2] Translation corpora are provided by: "Doosan Dong-A", "Neungyule Education", "Oxford University Press", "YBM Sisa.com", "Naver professional academic", "Web-crawl sentence", "English Hidden Card", "TIMES CORE" and "HarperCollins Publishers Ltd"

(2.36) ENG: Tax Time is **approaching**.
KOR: 세금 정산 시기가 다가옵니다
segeum jeongsan sigi-ga daga-obnida
Tax Calculation time-SUBJ close-come.POL
"Tax time is **coming closer**"

(*Source: Neungyule Education*)

According to an unpublished study by Cho (2009) evidence is found that Korean translators perform two strategies to circumvent the Moving Ego metaphor. The first strategy involves the replacement of motion verbs with their paraphrased meanings, as illustrated in examples (2.37 – 2.40).

(2.37) ENG: a child **reaching** adulthood.
KOR: 성년이 되어 가는 아이
seongnyeon-i doeeo ga-neun ai
adult-SUBJ become go-PROG child
"A child **becoming** an adult"

(*Source: YBM*)

(2.38) ENG: My son is **going through** a difficult phase.
KOR: 내 아들이 힘든 단계를 거치고 있다
nae adeul-i himdeun dangye-leul geochi-go iss-da
my son-SUBJ difficult phase-ACC suffer-PROG COP.PRS-PLA (=Plain)
"My son is **suffering** a difficult phase."

(*Source: Oxford University Press*)

(2.39) ENG: In the recession, our firm **went through** a bad time.
KOR: 우리 회사는 불경기 때 힘든 시기를 겪었다
uli hoesa-neun bulgyeonggi ttae himdeun sigi-leul gyeokk-eoss-da
our firm-SUBJ.FOC recession in hard time-ACC experience-PST-PLA
"In the recession, our company **experienced** a hard time."

(*Source: Oxford University Press*)

(2.40) ENG: When I **reach the end of the road** (bury me in a quiet place, near some trees.)
KOR: 내가 죽으면 ...
naega jugeu-myeon ...
I-SUBJ die-COND ...
"When I **die** ... "

(*Source: YBM*)

The examples (2.37 – 2.39) show that Moving Ego expressions are replaced by words of transformation ("become") and emotions ("suffer", "experience"). Example (2.40) which is beautifully poetic and metaphorical is simply reduced to "die". This example clarifies that the translator preferred to let the speaker simply "die" rather than to let him walk through time – towards the end of his life.

The second strategy involves the conversion of the Moving Ego metaphor to the Moving Time metaphor through the reversal of figure (ego) and ground (time landscape). The noun phrase "the Tenth Anniversary of our establishment" in (2.41) which forms the direct object in the original English sentence is moved to subject position and the former subject "we" is ommited. This phenomenon is called "Time-Model-Switching" (Cho, 2009).

(2.41) ENG: Since we are **approaching** the Tenth Anniversary of our establishment, ...
KOR: 창립 10주년 기념일이 다가오므로 ...
changlib 10 junyeon ginyeomil-i daga-omeulo ...
establishment 10 years anniversary-SUBJ close-come.as ...
"As the Tenth Anniversary of our establishment is **approaching**, ... "

In conclusion, translations of English Moving Ego sentences were either modified through paraphrase or Time-Model-Switching. It can be assumed that the Moving Ego metaphor is rejected in Korean and that Lakoff and Johnson (1999)'s prediction needs to be reconsidered. This matter will be taken up again in *Conclusion and Future Research* (Chapter 7) as the automatic detection of conceptual time metaphors might play an important role in natural language applications such as machine translation.

Chapter 3

Goals, Data and Challenges

3.1 Goals and Requirements

The goal of this research is to develop a program that detects and returns linguistic realizations of all three conceptual time metaphors in a given text. Sentences such as (3.1 – 3.4) are given as input to the program.

(3.1) "The time has come to make that choice, to act on what we know."

(3.2) "We're not going to go back to the days when insurance companies wrote the rules that let you languish without health care because you had a preexisting condition."

(3.3) "We're facing a time of trial and testing."

(3.4) "As I said when I announced this strategy, there will be more difficult days ahead."

At the end of the workflow the examples are returned as (3.5 – 3.8) by the CTM Detector.

(3.5) The [time: TIME] has [come] to make that choice, to act on what we know. - MOVING TIME

(3.6) We're not going to [go] back to the [days: TIME] when insurance companies wrote the rules that let you languish without health care because you had a preexisting condition. - MOVING EGO

(3.7) We're [facing] a [time: TIME] of trial and testing. - TIME ORIENTATION

(3.8) As I said when I announced this strategy, there will be more difficult [days: TIME] [ahead]. - TIME ORIENTATION

These results illustrate the successful recognition of different instances of conceptual time metaphors. The first example (3.5) contains a Moving Time metaphor, the Moving Ego metaphor is detected in example (3.6) and an expression of the Time Orientation metaphor is found in the last example (3.8). So the major goal was to develop a program that manages to automatically get from plain text input to a suitably tagged output. On the one hand, time and event nouns are enclosed in square brackets and tagged with the labels "TIME" or "EVENT". On the other hand, motion verbs and adverbs describing the time or event nouns are marked with square brackets, e.g. "[come]" (3.5), "[go]" (3.6)[1] and "[ahead]" (3.8). Information about the type of conceptual time metaphor is mentioned at the end of every line. These are either "MOVING TIME", "MOVING EGO" or "TIME ORIENTATION".

One of the motivations behind developing the CTM Detector was to lighten the researchers' workload of having to analyze an entire corpus manually. So it is important that the program can analyze arbitrary English language data without requiring the user to intervene during the different parsing and analysis steps. The entire program consists of a sequence of self-written scripts and several third party tools that will be introduced later. Filtering steps needed to be included into this processing chain in order to make the program as efficient as

[1] The progressive form "going" in the "going to"-construction (3.6) is not marked, as the construction expresses intentionality or prediction, rather than motion in time.

possible. The highest priority was placed on the quality of the output. Extensions and modifications of the program that raised the quantity and quality of the output were built into the program. In a nutshell, the program needs to be flexible, easy to use, efficient and accurate.

3.2 Data

Examples (3.1 – 3.4) are quotes by president Barack Obama taken from the CORpus of Political Speeches[2] (Guerini et al., 2008). This corpus is a collection of more than 3600 presidential speeches (about 7.9 millions words) and includes tags such as "APPLAUSE" and "LAUGHTER" that describe the audience's reactions. It was decided to use this corpus in the present research as the analysis of political discourse with respect to conceptual metaphors has a long tradition. This trend was also triggered by many studies by Lakoff who analyzed conceptual metaphors in various subdomains of political discourse. His study *Moral Politics: What Conservatives Know That Liberals Don't* (1996, 2005) describes the use of conceptual metaphors in domestic politics. One of his bold claims was that conservatives were more successful in convincing their listeners due to their more frequent use of conceptual metaphors. The liberals would lack the power of persuasiveness as they prefer to use facts and statistics. "*Metaphor and War: The Metaphor System Used to Justify War in the Gulf*" (Lakoff, 1992) illustrates the psychological effects that applications of conceptual metaphors can have on listeners: "Metaphors can kill. The discourse over whether we should go to war in the gulf is a panorama of metaphor.[...] The President says that the US is in the gulf to ''protect freedom, protect our future, and protect the innocent'', and that we must ''push Saddam Hussein back.'' Saddam is seen as Hitler. It is vital, literally vital, to understand just what role metaphorical thought is playing in bringing us to the brink of war." (Lakoff, 1992). Observations like these led to the conclusion that data from political discourse, especially in presidential speeches, could be a great resource for this project. It is rich of conceptual metaphors and I had high hopes that many time metaphors could be detected in this media.

The data in CORPS needed to be prepared in order to get them into the CTM Detector. First, all the metadata, e.g. the header and descriptions of the audience's feedback, were removed from all the speeches in the CORPS corpus and saved in another directory[3]. Those files that contain speech data from the same president were merged together to a single file[4]. The data used for training and testing is a merged collection of all the speeches given by the American president Barack Obama in the stretch of 2009 and 2010. It comprises about 43.600 sentences with 787.733 words and it served as input during the developmental stage of the CTM Detector.

3.3 Challenges

Many questions arose before and during the software development process. They were all related to linguistic questions and can be grouped together as **Semantic** and **Syntactic Challenges**.

3.3.1 Semantic Challenges

Time/Event Lexicon

One of the major challenges of programming the CTM Detector was the recognition of nouns that denote the lexical meaning of a time or event. So what the program needed was a lexicon containing nouns such as:

- **Time Nouns** day, week, holidays, christmas, easter[5]
- **Event Nouns** match, party, contest, election, tournament

A differentiation has to be made between time (e.g. "week") and event (e.g. "campaign") nouns. The motivation behind making this clear distinction is that the combination between Moving Ego and event nouns

[2]abbr. "CORPS"; Source: `http://hlt.fbk.eu/corps`
[3]FILES: "**corps-preprocessing.pl**" — LOCATION: "**./CORPS-Tools**")
[4]FILES: "**corps-text_merger.pl**" — LOCATION: "**./CORPS-Tools**")
[5]Case-insensitive

can cause erroneous output. The sentence "I am going to the match." is not an instance of a conceptual time metaphor as "match" is a metonym for "stadium". In these cases motion is happening in a spatial, and not in a temporal landscape. In order to make the program more accurate multi-token nouns had to be added to the lexicon as well:

- **Multi-Token Time/Event Nouns** new year, new year's eve

Motion Lexicon

The program needed a lexicon containing motion verbs. What seemed trivial at first sight turned out to be more complicated then expected. Not only was it important to determine which motion verbs could occur in the Moving Ego and Moving Time metaphors, but different aspects needed to be taken into consideration as well. All the word forms of the motion verbs needed to be included. And phrasal verbs which receive their motion sense only through co-occurrence had to be considered as well:

- **Motion Verbs used in Moving Ego/Time metaphors** go, come, approach, enter, exist
- **All Word Forms** go, went, gone, going
- **Phrasal Verbs** get close, draw near

Another interesting question was how the concept of the so-called "passage of time" is described in the Moving Time metaphor. Some relevant verbs are the following:

- **"Passage of Time"-Motion Verbs** fly, crawl, march

These groups illustrate that reasonable decisions needed to be made on which verbs were to be selected and added to the motion lexicon. The range of motion verbs used in conceptual time metaphors is limited as verbs such as "swim", "sprint" and "duckwalk" are probably not found. The motion lexicon needs to contain at least those verbs that are most frequently found in constructions of conceptual time metaphors.

Sight Lexicon

Sentences such as (2.5) and (2.6) showed that the ego can "look" forward and back within the Time Orientation metaphor. The ego can also "face difficult times" in the same way as facing another person. However, not all verbs of sight can be used within the Time Orientation metaphor. Sentences such as (3.9) and (3.10) reveal that "observing" and "watching" might be verbs of sight that are probably not found in the Time Orientation metaphor.

(3.9) ?We are observing great times.

(3.10) ?The team is watching the future.

A closer analysis has to be made on which verbs of sight are predominantly used in conceptual time metaphors. Based on this analysis, a sight lexicon needs to be created.

Preposition Lexicon

Last but not least, attention had to be paid on prepositions in prepositional or adverbial phrases occurring with motion verbs. The following two sentences illustrate the challenge:

1. We are arriving **in** a moment.
2. We are arriving **at** the moment of truth.

The Moving Ego metaphor occurs in the second, but not in the first sentence. The difference seems to lie in the prepositions that are used in both sentences. It does not necessarily mean that whenever there is a combination of "arrive"+"at"+[time|event] we have a Moving Ego metaphor ("We will arrive at the end of next month."), but more importantly, the preposition "in" does not seem to be used to create directionality, whereas "at" might do. A list of prepositions denoting directionality is therefore needed.

So in total, three different lexica (Time/Event, Verbs of Motion/Sight and Prepositions) are needed as essential components of the CTM Detector.

3.3.2 Syntactic Challenges

Questions concerning the necessity of a parser arose as well. One the one hand, one could try to find conceptual time metaphors without using any syntactic knowledge at all. On the other hand, the program could analyze the relations that exist within a sentence by running a parser.

No Parsing

The first possibility is to simply analyze strings of characters without taking linguistic meta-information into account. Sentences such as "The *deadline* is *coming* closer." could be detected by a regular expression such as (3.11)[6].

(3.11) *%TimeLex* (is |are)?*%MotionverbLex*

In theory, the regular expression would work if the variables "%TimeLex" (Time/Event lexicon) contained "deadline" and "%MotionverbLex" (Motion Verb lexicon) contained all word forms of the lexeme "come". An approach like this has the advantage that one could simply send several of these regexes through the entire input and hope that many sentences fit the regex patterns. However, linguistic phenomena such as syntactic movements are difficult to capture without any parsing component, as illustrated by examples (3.12 – 3.13).

(3.12) It's our **our childhood**, we want to return to.

(3.13) Those **times** we have gone through were difficult.

In the first example (3.12), the complement of the verb "return" is syntactically moved via **clefting**. If regular expressions were added to capture each and every linguistic phenomenon, the program would contain too many extremely complicated patterns. One can also assume that many regular expressions cannot even be formulated due to complexity reasons. The second case (3.13) is difficult to capture due to the use of an implicit **zero relative pronoun**. It is almost impossible to detect the conceptual time metaphor with a regular expression as the relative sentence is not explicitly marked with a relative pronoun ("Those times that we have gone through").

Despite the fact that an approach without parsing could be very fast, the limitations do overshadow the advantages as they have a negative effect on the quality and quantity of the output.

Parsing

The patterns we want to detect are of a semantic nature, so in principle the program needs functional specifications in forms of semantic roles, e.g. AGENT and GOAL, that specify the roles that participants have in a given situation. The program could detect conceptual time metaphors if it managed to answer the questions in (3.14 – 3.16).

(3.14) Is the time unit the AGENT?

(3.15) Is the AGENT moving?

(3.16) Is the time unit the GOAL?

[6]Represented as a pseudocode.

One the one hand, Moving Time metaphors could be detected by positive answers to (3.14) and (3.15). In sentence (3.17), "Christmas" is the AGENT argument of the predicate "approaching". This sentence combines positive answers to (3.14) and (3.15), and would be returned as an instance containing the Moving Time metaphor.

(3.17) Christmas is approaching.

On the other hand, time units labeled as GOAL (3.16) might indicate that the sentence contains a Moving Ego metaphor, as GOAL represents a target that an AGENT entity is moving towards to (3.18).

(3.18) We/AGENT are approaching Christmas/GOAL.

Sentences like these are distinguished from other sentences with temporal expressions, as time units are normally labeled as TIME (temporal placement) as in (3.19).

(3.19) We/AGENT are coming during Christmas/TIME.

These examples show why it might be desired to work with semantic roles. However, the fields of Semantic Role Labeling is still in the early stages and researchers working on programs that automatically label semantic roles still face many challenges: "To date, SRL systems have been shown to perform reasonably well in some controlled experiments, with F1 measures in the low 80s on standard test collections for English. Still, a number of important challenges exist for future research on SRL." (Màrquez et al., 2008). An alternative and, in practice, equally good solution is found in parsers that do not analyze sentences with semantic, but with grammatical relations such as "SUBJECT", "direct OBJECT" and "indirect Object". Normally the difference between these two types of relations becomes apparent through passivation of sentences. In both sentences "Sara hit John" and "John was hit by Sara", John is the patient and Sara the agent. In terms of grammatical relations, Sara is subject in the first but object in the second sentence. Despite this difference, one can say that working with grammatical relations represents an almost equally good solution, as passive constructions are almost never found in conceptual time metaphors, or would require a lot of creativity to produce one (3.20).

(3.20) We are being approached by difficult times.

In the end, the previously mentioned questions (3.14 – 3.16) are changed to (3.21 – 3.23).

(3.21) Is the time unit the SUBJECT?

(3.22) Is the SUBJECT moving?

(3.23) Is the time unit the Direct OBJECT?

Parsers that provide these grammatical categories return their output in forms of phrase or dependency structures. They vary according to the grammar that is used and the depth of information that is provided. The decision on which parser was used is described in the section *Syntactic Parser* (Section 4.3).

Chapter 4

CTM Detector: Preprocessing

This chapter will focus on the preprocessing components of the CTM Detector. This comprises the first four modules from "*Time/Event Tagger*" to "*Motion/Sight Tagger*" that are listed below. In these four steps the input is tagged, filtered and parsed. The final module "*CTM Extraction*" analyzes the fully preprocessed data and is responsible for extracting those sentences containing patterns that are typically found in expressions of conceptual time metaphors. This complex module is described in an extra chapter called "*CTM Detector: Extraction*" (Chapter 5).

1. **Time/Event Tagger**
2. **Time/Event Filter**
3. **Syntactic Parser**
4. **Motion/Sight Tagger**
5. **CTM Extraction** (Chapter 5)

The input data[1] has to be in plain text format and sentence borders do not have to be specifically marked, e.g. by newlines. The program executes sentence boundary recognition and splits sentences automatically in an earlier stage of the processing-flow.

4.1 Time/Event Tagger

The task of the Time/Event Tagger[2] is to identify and tag nouns that are either time or event nouns. Examples (4.1) and (4.2) contain the temporal nouns "week", "holiday", "festival of lights", "campaign" and "times".

(4.1) "And I think it's fitting that we begin this work in the week leading up to the holiday of Diwali the festival of lights when members of some of the world's greatest faiths celebrate the triumph of good over evil."

(4.2) "As I said in the campaign and as I've repeated many times as President the greatest generator of jobs in America is our private sector."

These sentences are transformed to (4.3) and (4.4).

(4.3) "And I think it's fitting that we begin this work in the time(week) leading up to the time(holiday) of Diwali the time(festival of lights) when members of some of the world's greatest faiths celebrate the triumph of good over evil."

(4.4) "As I said in the event(campaign) and as I've repeated many time(times) as President the greatest generator of jobs in America is our private sector."

[1] FILE: "**(input)**" — LOCATION: "**./CTM_Detector**"
[2] LOCATIONS: "**./CTM_Detector/1_TimeEventTagger_multiGram**" and "**./CTM_Detector/3_TimeEventTagger_uniGram**"

The program is not only meant to identify simple temporal nouns such as "week", "holiday", "campaign" and "times", but also multi-gram units, such as "festival of lights". These multi-token units are especially interesting if the temporal sense is not apparent from it's compositional units: "Tour de France" and "The Three Holy Hierarchs".

The various lexica with time nouns, event nouns and the different lengths are extracted from WordNet.

4.1.1 WordNet

WordNet[3] is a freely accessable, lexical database for English, containing semantic information about most English nouns, verbs, adjectives and adverbs. Words that share certain semantic features are grouped together to sets of synonyms which are called **synsets**. The synset "`{large,big,huge}`", which is enclosed in curly brackets, is just one of approximately 117,000 synsets. A pair of synsets can stand in different semantic relations:

Antonymy `{large, big, huge}` and `{small, little}`

Meronymy `{building,construction}` and `{elevator,lift}`

Hyponomy `{animal}` and `{dog}`

This database is a complex semantic network as synsets are interlinked with one another in multiple ways. Concepts lying closer to one another are more related than distant ones. This knowledge source is extremely valuable for the CMT Detector as it provides information about whether a noun has a time or event sense. Once WordNet is installed, specific information about a word ("Christmas") can be retrieved with commands such as (4.5) via the command-line interface.

(4.5) wn "Christmas" -hypen

The command consists of the abbreviation of WordNet **`wn`**, followed by the query word in quotation marks and closed by an option such as `-hypen`. This option is initialized with a dash, followed by the semantic relation Hypernyms (abbr. as **`hype`**) and the final letter **`n`** specifies that the target word is a noun. Once entered, WordNet returns the different senses of the search word, including a chain of synsets, as shown in Listing 4.1. Words which are in the same line are synonyms and form a synset together. The synset-levels are structured according to hyponymy relations and the researchers of the WordNet project determined the order by estimated frequency. Synsets at the lower level are the hypernyms, and synsets at the higher level are hyponyms. One of the synsets that is useful for the development of the time dictionary is "`{time period, period of time, period}`" (code line 8, 20 and 31 in Listing 4.1). Time nouns that occurred in conceptual time metaphors were best captured if they contained one of the three synsets (4.6 – 4.8).

(4.6) **`{time period, period of time, period}`**: e.g. vacation, period, holiday

(4.7) **`{point, point in time}`**: e.g. moment, date, beginning, end

(4.8) **`{time unit, unit of time}`**: e.g. minute, second, hour

Synset (4.6) captures time units that are mostly associated with the Duration Sense (described in Section 2.3). "Vacations" and "holidays" form a temporal interval initialized with an onset and closed with an offset. The second synset (4.7) fetches lexical entries that are conceptualized as specific points, and not intervals, in time. These are time nouns primarily bearing the Moment Sense (as defined in Section 2.3). The last synset (4.8) retrieves temporal concepts that represent "time-defining events" (described in Section 2.1).

The only synset that proved to be useful for capturing event nouns was (4.9).

(4.9) **`{social event}`**: e.g. political campaign, campaign, contest, competition

Attempts to capture other synsets such as "`{event}`" failed as too many irrelevant event nouns were matched as well, as illustrated in (4.10).

[3]Source: `http://wordnet.princeton.edu/`

(4.10) **{event}**: e.g. trouble, accident, stroke, fortuity

```
Synonyms/Hypernyms (Ordered by Estimated Frequency) of noun christmas

2 senses of christmas

Sense 1
Christmas, Christmastide, Christmastime, Yule, Yuletide, Noel
       => season
            => time period, period of time, period
                 => fundamental quantity, fundamental measure
                      => measure, quantity, amount
                           => abstraction, abstract entity
                                => entity

Sense 2
Christmas, Christmas Day, Xmas, Dec 25
       => legal holiday, national holiday, public holiday
            => holiday
                 => day
                      => calendar day, civil day
                           => time period, period of time, period
                                => fundamental quantity, fundamental measure
                                     => measure, quantity, amount
                                          => abstraction, abstract entity
                                               => entity
       => holy day of obligation
            => Christian holy day
                 => religious holiday, holy day
                      => holiday
                           => day
                                => calendar day, civil day
                                     => time period, period of time, period
                                          => fundamental quantity, fundamental measure
                                               => measure, quantity, amount
                                                    => abstraction, abstract entity
                                                         => entity
```

Listing 4.1: WordNet Entry of “Christmas”

4.1.2 Creation of Time and Event Lexica

The necessity of differentiating between time and event nouns was described in the section *Semantic Challenges* (Section 3.3.1). In order to implement this idea, two different groups of lexica are created. The first group contains only time nouns and consists of six lexica. Time nouns within these lexica have the same number of word n-grams and each noun occurs as singular, as well as plural form. The purpose of having six different lexica for time nouns is described later in the section *Time/Event Tagger: From Input to Output* (Section 4.1.3). The following lists are samples of time nouns that are stored in the six time lexica:

- **TimeLex1**: time, times, moment, moments, period, periods
- **TimeLex2**: arrival time, arrival times, departure time, departure times
- **TimeLex3**: All Saints’ Day, All Saints’ Days, Day of Atonement, Days of Atonement
- **TimeLex4**: Epiphany of Our Lord, Epiphanys of Our Lord
- **TimeLex5**: turn of the century, turns of the century
- **TimeLex6**: Struggle for Freedom and Democracy Day, Struggle for Freedom and Democracy Days

The event nouns are created in the same way:

- **EventLex1**: concert, concerts, dogfight, dogfights, funeral, funerals
- **EventLex2**: auto race, auto races, athletic contest, athletic contests
- **EventLex3**: campaign for governor, campaigns for governor
- **EventLex4**: (empty)
- **EventLex5**: Iditarod Trail Dog Sled Race, Iditarod Trail Dog Sled Races

Each lexicon is stored separately as hash tables in files called "TimeLex1", "TimeLex2", etc. The number of time and event dictionaries is determined by the longest entries that were found in WordNet. The longest time noun was a 6-gram sequence and the longest event noun consisted of five units. EventLex4 is created even though it is empty due to expandability reasons: the program must be able to deal with 4-gram event nouns that are e.g. added later.

These lexica are created with perl scripts for time[4] and event[5]. In general, the functions of these scripts are to extract all nouns from the WordNet noun database, to look them up individually in WordNet, to create their plural forms and to store them if they contain a time or event synset. These four steps are described in more detail now.

The list of nouns is retrieved from the WordNet noun database "`data.noun`"[6]. This file contains nouns (e.g.: "arrival_time", "time_of_arrival"), encyclopedic information and numeric IDs that index the entries and build references to other related ones:

```
15180934 28 n 02 arrival_time 0 time_of_arrival 0 001 @ 15180528 n 0000 | the
time at which a public conveyance is scheduled to arrive at a given destination
15181094 28 n 02 departure_time 0 time_of_departure 0 002 @ 15180528 n 0000  15181282
n 0000 | the time at which a public conveyance is scheduled to depart from a given
point of origin
15181282 28 n 02 checkout 0 checkout_time 0 002 @ 15181094 n 0000 + 00966492 v 0101
| the latest time for vacating a hotel room; ``the checkout here is 12 noon"
```

In total, 119000 nouns are retrieved from this database with the help of a regular expression and saved in a temporary hash dictionary (%lexicon_entries). Each entry is then looked up in WordNet with the commands in Listing 4.2.

```
# Set Environment Variable: WORDNET
$ENV{'PATH'} = '/usr/local/Cellar/wordnet/3.0/bin';

# Go Through Every Lexicon Entry
foreach (keys %lexicon_entries) {
  print $_."\n";

  my $wnSenses = `wn ``$_`` -hypen`;
...
```

Listing 4.2: WordNet Environment Setup

The semantic information with all the synsets (as in example 4.1) are retrieved from WordNet and saved in the local variable $wnSenses. The script runs regular expressions through the chains of synsets and checks if it contains one of the three time synsets or the "`{social event}`" synset. If it does, then the plural form of the noun is generated through the perl module called `Lingua::EN::Inflect`[7]. The simple command "`PL($_)`"

[4]FILE: "**extraction_time.pl**" — LOCATION: "**./Lexica/Time**"

[5]FILE: "**extraction_event.pl**" — LOCATION: "**./Lexica/Event**"

[6]WordNet 3.1 Database files: `http://wordnetcode.princeton.edu/wn3.1.dict.tar.gz` (26.11.2012)

[7]LOCATION: "**./Lexica/modules**" — Source: `http://search.cpan.org/~dconway/Lingua-EN-Inflect-1.895/lib/Lingua/EN/Inflect.pm` (27.11.2012)

returns the plural form. The length of the singular and plural form of the time/event noun is determined and depending on this, the nouns are saved in one of the eleven lexica. The factor if the time or event sense was detected in the first, second or third (or later) sense provides valuable information about how strongly the noun is associated with temporal concepts. For this reason the numbers 1, 2 or 3 are assigned as values of the respective hash keys, as shown in Listing 4.3.

```
$VAR1 = 'pubescences';
$VAR2 = '1';
$VAR3 = 'MArch';
$VAR4 = '1';
[...]
$VAR117 = 'endings';
$VAR118 = '3';
[...]
$VAR541 = 'climax';
$VAR542 = '2';
```

Listing 4.3: Data Dumper: TimeLex1

In addition, another source was taken to expand the time lexica. The English Wikipedia entry "List of holidays by country"[8] contains links to lists of national holidays of 192 countries. These lists were manually copy-pasted to an excel file, sorted and handed over to a perl script[9] that returns all holidays that occurred at least two times. 219 additional time nouns were added to the lexica through this procedure.

The statistics in Table 4.1 and Table 4.2 reveal how many entries were stored in the time and event lexica in the end.

2381	= Sense 1
86	= Sense 2
30	>= Sense 3
2497	Total

Table 4.1: Statistics – Time Nouns

645	= Sense 1
41	= Sense 2
14	>= Sense 3
700	Total

Table 4.2: Statistics – Event Nouns

Almost 3200 time and event nouns were collected in total and 95% of them are primarily associated with either time or event (Sense 1). In fact, the other 5% were kept as well as they included time nouns such as "generation" and "youth" that could be found in time metaphors.

4.1.3 Time/Event Tagger: From Input to Output

The advantage of having several lexica with nouns of different n-gram length is that the Time/Event tagger can simply send regular expressions through the entire input data and search for time nouns in the following order: 6-gram, 5-gram, 4-gram, 3-gram 2-gram and 1-gram time nouns, and search after event nouns in the same manner afterwards. A sentence such as "New Year's Eve is getting closer." could be transformed by TimeLex3 to "time(New Year's Eve) is getting closer.", and the subsequent two dictionaries (TimeLex2 and TimeLex1) would not detect anything as the time reference is already semantically tagged. But problems would occur with zero derivation words such as "run" which could be both, event noun and verb. In order to avoid erroneous outputs, such as "I will event(run) the event(marathon) at time(New Year's Eve)." the program needs to check if the word that is matched has the part-of-speech "noun". For this reason an intermediate step, a part-of-speech tagger is added:

1. TIME Lexicon: 6-gram, 5-gram, 4-gram, 3-gram, 2-gram
2. EVENT Lexicon: 5-gram, 4-gram, 3-gram, 2-gram
3. **POS Tagger: CRFTagger**

[8] Source: http://en.wikipedia.org/wiki/List_of_holidays_by_country

[9] FILE: "**extract_holidays.pl**" — LOCATION: "**./Lexica/Time/NationalHolidays**"

4. TIME Lexicon: 1-gram

5. EVENT Lexicon: 1-gram

The java-based **CRFTagger**[10] (Phan, 2006) is a conditional random fields tagger for English. It is chosen due to its high accuracy (97%) and its fast processing speed (approx. 500 sentences/second). The following step-to-step overview (Table 4.3) illustrates how the previous sentence "I will run the marathon at New Year's Eve." is transformed in five steps. The underlined components highlight which processes were activated (left column) and which changes were consequently made with the input (right column).

INPUT	**I will run the marathon at New Year's Eve.**
1. TIME 6-5-4-3-2	I will run the marathon at time(NewˆYear'sˆEve).
2. EVENT 5-4-3-2	I will run the marathon at time(NewˆYear'sˆEve).
3. CRFTagger	I/PRP will/MD run/VB the/DT marathon/NN at/IN time/NN (/(NewˆYear'sˆEve/NN)/) ./.
(pos/ˆ removal)	I/PRP will/MD run/VB the/DT marathon/NN at/IN time(New Year's Eve) ./.
4. TIME 1	I/PRP will/MD run/VB the/DT marathon/NN at/IN time(New Year's Eve) ./.
5. Event 1	I/PRP will/MD run/VB the/DT event(marathon) at/IN time(New Year's Eve) ./.
OUTPUT	**I will run the event(marathon) at time(New Year's Eve).**

Table 4.3: Time/Event Tagger

Spaces between multigram nouns are replaced with carets to keep the units together before they are processed by the CRFTagger. The carets are removed with a regular expression after the part-of-speech tagging. In the end, the Time/Event Tagger has correctly identified two instances of temporal concepts ("marathon" and "New Year's Eve") and managed to identify "run" correctly as verb.

4.2 Time/Event Filter

The task of the Time/Event Filter[11] is to significantly reduce the size of the input text by removing those sentences that do not contain temporal nouns. The training file containing all the Obama speeches contains about 43.600 sentences. After removing irrelevant sentences, the corpus is reduced to 11023 sentences, which is approximately 75% smaller than the original. The advantages of having this smaller file are **practicality** and **efficiency**. Some researchers who prefer to analyze corpora manually (Cameron and Deignan, 2003; Charteris-Black, 2004) could use this filtered text data, instead of the entire corpus. All the sentences in the data represent potential candidates of conceptual time metaphors and the fact that temporal nouns are tagged makes analysis much easier and much more efficient. But more importantly the processing time of the subsequent step, the syntactic parser, is reduced significantly. It is ensured that time is not wasted with processing sentences that do not contain temporal references.

4.3 Syntactic Parser

In the section *Syntactic Challenges* (Section 3.3.2) the advantages of embedding and not embedding a parser is briefly discussed. In the end, the decision in favor of building in a parser was made. The range and variety of parsers is very wide. Currently, there are two major types of representing the structure of a sentence.

The first option is to employ tools that create **phrase structures** or **constituent structures** as defined by a "**context-free grammar**" (Chomsky, 1956). The original context-free grammar representation however does not provide any functional specifications such as grammatical relations. For this reason, only parsers that return **enriched phrase structure parses** e.g. the "Lexical Functional Grammar" (Kaplan and Bresnan, 1995) and the "Head-Driven Phrase Structure Grammar" (HPSG) (Pollard and Sag, 1994) came into question. Even

[10]LOCATION: "**./CTM_Detector/2_CRFTagger**" — Source: `http://crftagger.sourceforge.net/`
[11]LOCATION: "**./CTM_Detector/4_TimeEvent_Filter**"

though these extended phrase structure grammars may provide the necessary grammatical relations, several arguments speak against them. The creation of the phrase structures with all the attribute value matrices and the extraction of relevant temporal nouns within these complex units is time-consuming. Most sentences contain syntactically ambiguous segments that have to be manually resolved. These parsers might also have troubles working with the output of the Time/Event Tagger as it comes with temporal nouns enclosed within "time()" or "event()" tags.

An alternative way to determine and represent sentence structure is **dependency parsing**. Dependencies are syntactic constructions that describe *asymmetrical binary relations* between **lexical elements**. Phrasal nodes do not exist in this framework, as syntactic connections are only given between one word and another. These relations on the word level describe the dependency between a "**head**" and a "**dependent**". This framework originates in Tesnière's study "Éléments de syntaxe structurale" (Tesnière, 1959), in which these two terms are introduced as "**régissant**" and "**subordonné**" respectively. All the dependency parsers vary with respect to how the dependencies between two units are determined (Zwicky, 1985; Hudson, 1987) and how these relations are specified.

Even though both approaches have their own strengths and advantages the decision was made to work with dependency structures. Dependency grammars provide valuable information about predicate-argument structures such as subjects and objects directly, without all the attribute value matrices that are components of enriched phrase structure grammars. Parsers which label the dependencies with these **grammatical relations** are referred to as "Typed Dependency Parsers". With this information the questions of "*whether the time unit is the subject*", "*whether the subject is moving*" and "*whether the time unit is the direct object*" can be quickly answered.

4.3.1 Stanford Parser

An adequate solution for solving the syntactic challenges is found in the **Stanford Parser**[12]. It contains, inter alia, a high-accuracy **unlexicalized probabilistic context-free grammar parser** (PCFG), developed by Klein and Manning (2003), and a **typed dependency parser** (de Marneffe et al., 2006). Both components are required as the dependency structure is directly extracted from the phrase structure parses via rules. The Stanford Parser returns the sentence structures (4.11) for the input "We are approaching New Year's Eve".

(4.11)
```
(ROOT
  (S
    (NP (PRP We))
    (VP (VBP are)
      (VP (VBG approaching)
        (NP
          (NP (NNP New) (NNP Year) (POS 's))
          (NNP Eve))))
    (. .)))
```

```
nsubj(approaching-3, We-1)
aux(approaching-3, are-2)
root(ROOT-0, approaching-3)
nn(Year-5, New-4)
poss(Eve-7, Year-5)
dobj(approaching-3, Eve-7)
```

The first half of the output represents the phrase structure parse. This is used as input for the dependency parser which detects dependencies between words and describes these relations as precisely as possible. Each line in the second half of the output represents a single dependency. Every dependency consists of a tag for the grammatical relation, the head and the dependent. In the first example, the dependent "**We**" stands in the grammatical relation **nominal subject** (abbreviated as "nsubj") to its head "**approaching**". The developers of

[12]LOCATION: "**./CTM_Detector/5_StanfordParser**" — Source: http://nlp.stanford.edu/software/lex-parser.shtml

this dependency parser collected 48 grammatical relations in total. These relations are organized hierarchically (de Marneffe et al., 2006), as showsn in Table 4.4.

dep - dependent
- aux - auxiliary
  - auxpass - passive auxiliary
  - cop - copula
- conj - conjunct
- cc - coordination
- arg - argument
  - subj - subject
    - nsubj - nominal subject
      - nsubjpass - passive nominal subject
    - csubj - clausal subject
  - comp - complement
    - obj - object
      - dobj - direct object
      - iobj - indirect object
      - pobj - object of preposition
    - attr - attributive
    - ccomp - clausal complement with internal subject
    - xcomp - clausal complement with external subject
    - compl - complementizer
    - mark - marker (word introducing an advcl)
    - rel - relative (word introducing a rcmod)
    - acomp - adjectival complement
      - agent - agent
- ref - referent
- expl - expletive (expletive there)
- mod - modifier
  - advcl - adverbial clause modifier
  - purpcl - purpose clause modifier
  - tmod - temporal modifier
  - rcmod - relative clause modifier
  - amod - adjectival modifier
  - infmod - infinitival modifier
  - partmod - participial modifier
  - num - numeric modifier
  - number - element of compound number
  - appos - appositional modifier
  - nn - noun compound modifier
  - abbrev - abbreviation modifier
  - advmod - adverbial modifier
    - neg - negation modifier
  - poss - possession modifier
  - possessive - possessive modifier ('s)
  - prt - phrasal verb particle
  - det - determiner
  - prep - prepositional modifier
- sdep - semantic dependent
  - xsubj - controlling subject

Table 4.4: Stanford Parser: Dependency Hierarchy

Most of them derive from the studies by Lin (1995) and Carroll et al. (1999). But some of them were added: "we have introduced a number of extensions and refinements to facilitate use in applications. Many NP-internal relations play a very minor role in theoretically motivated frameworks, but are an inherent part of corpus texts and can be critical in real-world applications. Therefore, besides the commonest grammatical relations for NPs (amod - adjective modifier, rcmod - relative clause modifier, det - determiner, partmod - participial modifier, infmod - infinitival modifier, prep - prepositional modifier), our hierarchy includes the following grammatical relations: appos (appositive modifier), nn (noun compound), num (numeric modifier), number (element of compound number) and abbrev (abbreviation)." (de Marneffe et al., 2006). This seemingly small extension is what makes this tool so powerful compared to other dependency parsers. The capability to handle appositive modifiers (abbreviated as "appos") creates the possibility to feed the parser with sentences that were previously modified by the Time/Event Tagger, as illustrated in (4.12) and (4.13).

(4.12) **We are approaching New Year's Eve**

nsubj(approaching-3, We-1)
aux(approaching-3, are-2)
root(ROOT-0, approaching-3)
nn(Year-5, New-4)
poss(Eve-7, Year-5)
dobj(approaching-3, Eve-7)

(4.13) **We are approaching time(New Year's Eve)**

nsubj(approaching-3, We-1)
aux(approaching-3, are-2)
root(ROOT-0, approaching-3)
dobj(approaching-3, time-4)
nn(Year-7, New-6)
poss(Eve-9, Year-7)
appos(time-4, Eve-9)

These two columns show how the parser analyzes the untagged and the tagged versions of the same sentence. In the tagged version (4.13), "Eve", which is the head of "New Year's Eve", is labeled as an appositive modifier of "time", and "time" successfully replaces "Eve" as dependent in the direct object dependency. This explains as well why temporal nouns such as "New Year's Eve" are placed in brackets right behind "time", and not the other way around. By embedding the "time" or "event" tags into the sentence, dependencies such as "dobj(approaching-3, time-4)" arise, which represent very potential candidates of Moving Ego metaphors. It is these kinds of dependencies that the CTM Detector will try to extract.

4.4 Motion/Sight Tagger

The next step is to go through the entire parse chain and to tag those verbs that are normally associated with locomotion or sight. This is where the motion and sight lexicon comes into play.

Motion Verbs

First attempts to create this lexicon by automatically extracting motion verbs from WordNet failed. The extraction of verbs that contain synsets, e.g. "`{travel, go, move, locomote}`", resulted in having many motion verbs but the majority of them were unusable as they were not applicable to the system of the three conceptual time metaphors. For example, verbs denoting motion in a vertical direction such as "climb", "rise", "mount", "descend" or "fall" are not used.

In addition, certain verbs are only used in the Moving Time, but not in the Moving Ego metaphor, and vice versa. Verbs that normally express motion in a spatial *room*, such as "enter", "step", "exit", "leave" and "escape" are only found in Moving Ego constructions (4.14 – 4.18).

(4.14) **enter** "high school students before they enter into adulthood" (COCA)[13]

(4.15) **step** "Step into summer with CLOGS." (COCA)

(4.16) **exit** "picture of not only the era which some people think we are exiting" (COCA)

(4.17) **leave** "My family had never really left the 19th century" (COCA)

(4.18) **escape** "He ends by describing his escape from the time period" (COCA)

The direct objects in these sentences have something in common. They all bear the Duration Sense. These time units are cognitively modeled as rooms that can be entered, gone through and existed in the Moving Ego metaphor. Motion of a time unit into, through and out of another temporal unit is not possible in the Moving Time model. The temporal sequence of events (e.g. Christmas ahead of New Year's Eve) is firmly determined and the change of this order is impossible.

Even though time "flows like a river", flowing or swimming motion is unlikely to be found in the Moving Ego metaphor. Time "flies" but one does not fly towards Christmas. Whenever it's boring time seems to "drag" or "crawl" by. But a person does not drag or crawl towards New Year's Eve. The major point is that verbs that describe the motion of time, are in many cases not used to describe the way the ego travels through time. Finding linguistic or cultural explanations for these differences is not inside the focal point of this paper. But a possible explanation for these differences might be that the concept of **path** is a component of the Moving Ego, but not of the Moving Time metaphor. The ego must utilize this path and move on through time like an ordinary traveler.

In addition, some verbs can be used in both time metaphors, as shown in (4.19 – 4.22)

(4.19) **approach – Moving Time** "The time was approaching when I" (COCA)

(4.20) **approach – Moving Ego** "we approach the time for trial." (COCA)

(4.21) **get close – Moving Time** "as the trial was getting closer" (COCA)

(4.22) **get close – Moving Ego** "it gets close to the harvest time" (COCA)

Both transitive verbs only denote unidirectional movement of one entity towards a target and the movement is not associated with temporal compression (e.g. time flies) or protraction (e.g. time drags/crawls).

All these observations led to the conclusion that the Motion Verb lexicon needed the following two features. The motion verbs must be stored with specific information on whether they can occur a) only in the Moving Time metaphor b) only in the Moving Ego metaphor or c) in both of them. The lists of motion verbs that are used in each metaphor model needed to be collected manually. The search engine of the Corpus of Contemporary American English provides a concordancer that enables to search after collocations between directional prepositions and verbs. Figure 4.1 shows how motion verbs were collected. The node of the search pattern is a group of prepositions, separated by alternation characters, that could be used as directional prepositions (example 4.23).

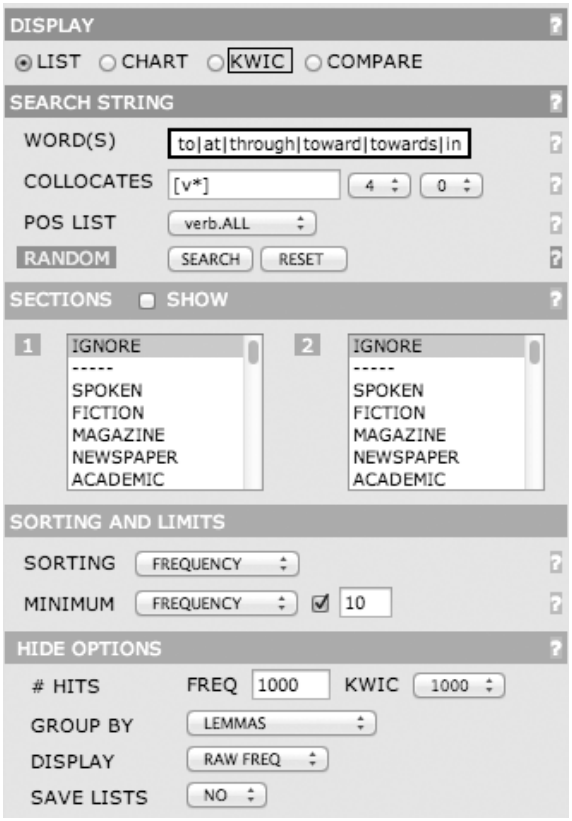
(4.23) "to|at|through|toward|towards|in|across|around|from|into|onto|out of"

With these settings, COCA returns a sorted list of verb lemmas that occurred at least 10 times in the span of -4,0. Table 4.5 shows a list of the 20 most frequent verbs. Verbs that were found in conceptual time metaphors were extracted by going through the list manually. The infinitive forms of the verbs that occurred either in the Moving Time and Moving Ego metaphors were saved in separate files[14], as shown in (4.24) for the Moving Time and in (4.25) for the Moving Ego metaphor.

[13] All these examples are taken from the **Corpus of Contemporary American English**, Source: `http://corpus.byu.edu/coca/`

[14] FILES: "**base_movingTime**", "**base_movingEgo**" — LOCATION: "**./Lexica/Verbs**"

Figure 4.1: Search Pattern for Motion Verbs



(4.24) **base_movingTime**
approach, arrive, come, crawl, drag, fly, follow, march, move, near, precede, reach, run, rush, pass, speed

(4.25) **base_movingEgo**
approach, enter, escape, exit, flee, go, head, near, reach, return, travel, walk, wander

A perl script[15] loads both lists and extracts all the present tense, past tense, participle and the progressive forms from an English verb inflection database[16] from the Nodebox English Linguistics Library[17] which contains word forms of 8567 verbs. The relevant word forms are saved in a hash variable. If the motion verb occurs only in the Moving Ego metaphor then it receives the value "motion_mE" (= Moving Ego). The value "motion_mT" (= Moving Time) is given if the verb is only used in the Moving Time metaphor. Motion verbs that can occur in both metaphors get the value "motion_mET" (= Moving Ego/Time). The advantage of having this extra script for creating the Motion Verb lexicon is extensibility. The user can easily add missing motion verbs to the lexicon by simply adding infinitive forms to the verb lists and running the perl script.

In the current state, the lexicon file "VerbLex"[18] is returned with the hash structure illustrated in Listing 4.4.

[15]FILE: "**extraction_verbs.pl**" — LOCATION: "**./Lexica/Verbs**"
[16]FILE: "**verb.txt**" — LOCATION: "**./Lexica/Verbs**"
[17]Source: `http://nodebox.net/code/index.php/Linguistics`)
[18]FILE: "**VerbLex**" — LOCATION: "**./CTM_Detector/6_MotionSightTagger**"

1	[BE]	3388525
2	[HAVE]	1356526
3	[GO]	786311
4	[WANT]	490037
5	[DO]	410245
6	[COME]	302327
7	[GET]	301915
8	[TRY]	294266
9	[LOOK]	258368
10	[WILL]	236997
11	[WOULD]	229370
12	[NEED]	190624
13	[TAKE]	175529
14	[SAY]	173293
15	[CAN]	169653
16	[SEEM]	155850
17	[USE]	154435
18	[MAKE]	131053
19	[BEGIN]	114861
20	[TURN]	111740

Table 4.5: Frequency List: 20 Most Frequent Verbs Co-occurring with Directional Prepositions

```
...
$VAR3 = 'fly';
$VAR4 = 'motion_mT';
$VAR5 = 'come';
$VAR6 = 'motion_mET';
$VAR7 = 'arriving';
$VAR8 = 'motion_mT';
$VAR9 = 'headed';
$VAR10 = 'motion_mE';
...
```

Listing 4.4: DataDumper: VerbLex (1)

With the help of the lexicon the Motion Verb Tagger[19] can process the entire dependency parses and label motion verbs, as illustrated in input (4.26) and output (4.27).

(4.26) **Before Motion Verb Tagging**

nsubj(approaching-3, We-1)
aux(approaching-3, are-2)
root(ROOT-0, approaching-3)
dobj(approaching-3, time-4)
nn(Year-7, New-6)
poss(Eve-9, Year-7)
appos(time-4, Eve-9)

[19]LOCATION: "**./CTM_Detector/6_MotionSightTagger**"

(4.27) **After Motion Verb Tagging**

nsubj(motion_mET:approaching-3, We-1)
aux(motion_mET:approaching-3, are-2)
root(ROOT-0, motion_mET:approaching-3)
dobj(motion_mET:approaching-3, time-4)
nn(Year-7, New-6)
poss(Eve-9, Year-7)
appos(time-4, Eve-9)

The motion verb "approach" is found in both conceptual time metaphors. The different word forms of this lexeme ("approach", "approaches", "approached", "approaching") were thus saved in the Motion Verb lexicon with the value "motion_mET". The tagger scans the parse, matches the motion verb "approaching" and adds the information that is underlined in (4.27).

Sight Verbs

The lexicon "VerbLex" is extended with so-called verbs of sight that are found in the Time Orientation metaphor. The only verbs that are taken into consideration are "look" and "face". They are loaded from the file "base_timeOrientation", all the word forms are looked up and added to the lexicon as well. During the developmental stage, other verbs such as "see" were also included in the lexicon but the precision rate dropped immensely as tests on the Obama speeches showed that only one out of 20 returned instances was actually correct. The only correct detection is given in example (4.28).

(4.28) ... you will [see] a [time: TIME] in which we put a stop to discrimination against gays and lesbians whether in the office or ... - TIME ORIENTATION

Two false positive cases of the combination between "see" and a time or event object are given in (4.29) and (4.30).

(4.29) We [saw] this [month: TIME], for the first [time: TIME] in three [years: TIME], manufacturing actually ticking up, consumer confidence is up. - TIME ORIENTATION

(4.30) Now, as we're about to [see] this [evening: TIME], there's nothing quite like the power and the passion of Broadway music. - TIME ORIENTATION

The challenge lies in the fact that the temporal noun phrases "this month" and "this evening" follow the verb "see" directly, which causes the grammatical relation between the verb and the noun to be incorrectly described as direct object. For this reason, the verb "see" is not included in the final version of the verb lexicon.

All the word forms of these two lexemes "look" and "face" are given the value "sight_TO" which abbreviates "Time Orientation". Listing 4.5 shows an extract of the final version of VerbLex which contains, inter alia, the word forms "faces" and "looks".

```
...
$VAR103 = 'faces';
$VAR104 = 'sight_TO';
$VAR105 = 'moved';
$VAR106 = 'motion_mT';
$VAR107 = 'looks';
$VAR108 = 'sight_TO';
$VAR109 = 'wander';
$VAR110 = 'motion_mE';
...
```

Listing 4.5: DataDumper: VerbLex (2)

Dependencies that included word forms of either "see" or "face" are labeled with the "sight_TO" tag. The labeling is illustrated with the dependencies of the sentence "Some people haven't faced their moment yet." in examples (4.31) and (4.32).

(4.31) **Before Verb Of Sight Tagging**

det(people-2, Some-1)
nsubj(faced-5, people-2)
aux(faced-5, have-3)
neg(faced-5, n't-4)
root(ROOT-0, faced-5)
poss(time-7, their-6)
dobj(faced-5, time-7)
appos(time-7, moment-9)
advmod(faced-5, yet-11)

(4.32) **After Verb Of Sight Tagging**

det(people-2, Some-1)
nsubj(sight_TO:faced-5, people-2)
aux(sight_TO:faced-5, have-3)
neg(sight_TO:faced-5, n't-4)
root(ROOT-0, sight_TO:faced-5)
poss(time-7, their-6)
dobj(sight_TO:faced-5, time-7)
appos(time-7, moment-9)
advmod(sight_TO:faced-5, yet-11)

The dependency parses are now ready to be searched for conceptual time metaphors. After all preprocessing steps described in this chapter, the input data is suitably parsed and tagged. The dependencies provide information about verbs of motion and sight, and whether the subject or object is a temporal noun, e.g. "dobj(motion_mET:approaching-3, time-4)". These chains of parses are handed over to the Conceptual Time Metaphor Extractor which is described in the following chapter.

Chapter 5

CTM Detector: Extraction

The task of the CTM-Extractor[1] is to detect dependency patterns that are typically found in linguistic realizations of conceptual time metaphors. Various regular expressions that capture Moving Time, Moving Ego and Time Orientation dependencies are sent through the parse. If a regular expression matches, the entire sentence is returned with tags and labels, as described in the section "*Goals and Requirements*" (Section 3.1).

At the first stage of this implementation it was expected that it would be sufficient to search after what will be referred to as "**primary dependencies**": dependencies that are most commonly found in each type of conceptual time metaphor. Most Moving Time metaphors might be detected with subject dependencies between a time noun and a motion verb (example 5.1). The detection of an object dependency between a motion verb a time noun might return most of the Moving Ego constructions (example 5.2). It was expected that Time Orientation metaphors could be all captured through dependencies between a verb of sight and a temporal noun (example 5.3), or through combinations of specific adverbs ("ahead"/"behind") and a time noun (example 5.4). However, the evaluation of the primary dependency rules showed that many sentences could not be retrieved and that further extensions and modifications were necessary.

(5.1) nsubj(motion_mT:came-8, time-4)

(5.2) dobj(motion_mET:approaching-3, time-4)

(5.3) dobj(sight_TO:facing-3, time-5)

(5.4) npadvmod(ahead-8, time-4)

This chapter is divided into three parts. First, the primary dependencies rules are described in the section *Primary Dependency Extraction* (Section 5.1). The so-called gold standard database which is used for evaluating these rules is introduced in the subsequent section *Gold Standard Database* (Section 5.2). The modifications and extensions which lead to the final version of the CTM Detector are described in the section *Extensions and Modifications* (Section 5.3).

5.1 Primary Dependency Extraction

5.1.1 Moving Time Dependency

The dependency relation between a temporal subject and a motion verb is crucial for detecting Moving Time metaphors. The dependency parse of the subordinate clause in "And when the moment came, they did what they were trained to do." is given in (5.5).

[1]LOCATION: "./CTM_Detector/7_CTM-Extraction"

(5.5) **And when the moment came** ...

cc(did-11, And-1)
advmod(motion_mT:came-8, when-2)
det(time-4, the-3)
nsubj(motion_mT:came-8, time-4)
appos(time-4, moment-6)
advcl(did-11, motion_mT:came-8)
...

Regular expressions (abbr. "regex") are needed in order to detect dependencies. The regular expressions that will capture them are perl regexes embedded in the match operator "m//". The letter "m" indicates that a string has to be *matched* and the slashes act as delimiters. The primary dependency of the previous example (5.5) is captured with the regular expression in Listing (5.1).

```
m/nsubj\((motion_mE?T:(.+?)-\d+), (time|event)-\d+\)/
```

Listing 5.1: RegEx: Moving Time – Primary Dependency

However, even though this simple regular expression managed to return all sentences containing this dependency, first tests showed that more than 50% of the output was incorrect. Among them was also the sentence: "And last year he approached Senator Kennedy to share his ideas about service". A dependency is mistakenly created between "year" and "approached", as illustrated in (5.6).

(5.6) **And last year he approached Senator Kennedy** ...

cc(motion_mET:approached-8, And-1)
amod(time-3, last-2)
nsubj(motion_mET:approached-8, time-3)
dep(motion_mET:approached-8, year-5)
nsubj(motion_mET:approached-8, he-7)
root(ROOT-0, motion_mET:approached-8)
...

Another wrong sentence was (5.7).

(5.7) **Most days, they'd come home.**

amod(time-2, Most-1)
nsubj(motion_mET:come-9, time-2)
dep(motion_mET:come-9, days-4)
nsubj(motion_mET:come-9, they-7)
aux(motion_mET:come-9, 'd-8)
root(ROOT-0, motion_mET:come-9)
dobj(motion_mET:come-9, home-10)

In both cases the temporal adjuncts ("last year" and "Most days") are topicalized and mistakenly classified as subjects of the sentences. Weaknesses of the Stanford Tagger caused these and other problems. However, one of the characteristics of the Moving Time metaphor is that it never comes with another non-temporal subject or object. Nobody would utter sentences such as "Christmas is getting closer to us" or "President Obama and Christmas are getting closer". Analysis has shown that the essential components of the typical Moving Time sentence are a temporal subject and a motion verb. The last two dependency parses (5.6) and (5.7) contained another non-temporal subject "nsubj(motion_mET:approached-8, he-7)" and a superfluous object "dobj(motion_mET:come-9, home-10)" (highlighted in italics). The addition of a simple conditional rule solved these problems. The regular expression in Listing 5.2 contains the negative lookahead "((?!(time|event))" which prohibits the existence of a subject, object and adjunct dependency between the motion verb and another non-temporal noun.

```
m/((dobj|nsubj|prep_.+?)\($nsubjMotion, ((?!(time|event))))/
```

Listing 5.2: Moving Time – Conditional Rule

The pseudo-code (5.8) illustrates how the conditional regular expression is embedded into the rule of extracting the Moving Time primary dependency.

(5.8) *If the program matches nsubj(motion verb, time):*
- If there is a dependency between the same motion verb and another non-temporal subject/object: Ignore sentence!
- Else: Return the sentence as Moving Time metaphor!

Fortunately the conditional rule (5.8) removed about 90% of the wrong sentences without notably reducing the number of correct results. The examples (5.9 – 5.12) illustrate samples of the output.

(5.9) The [time: TIME] has [come] to make that choice, to act on what we know. - MOVING TIME

(5.10) In the [weeks: TIME] and [months: TIME] that [followed], he battled to recover from traumatic brain injury. - MOVING TIME

(5.11) I mean, the [day: TIME] has [passed] when I expected this to be a full partnership. - MOVING TIME - MOVING TIME

(5.12) He invited them sailing, played with their children, and would write each family a letter whenever the [anniversary: TIME] of that terrible [day: TIME] [came] along.' - MOVING TIME

5.1.2 Moving Ego Dependency

In the Moving Ego metaphor the primary dependency is given between a motion verb and a temporal object or adjunct. This relation can be expressed either with preposition or without (the latter is labeled "dobj").

- **prep_to**(motion verb, time)
- **prep_at**(motion verb, time)
- **prep_through**(motion verb, time)
- **prep_toward(s)**(motion verb, time)
- **prep_into**(motion verb, time)
- **prep_from**(motion verb, time)
- **dobj**(motion verb, time)

All these prepositions are used to express directionality and saved in the variable "$prepositions" (Listing 5.3).

```
my $prepositions = ''to|at|through|towards?|into|from";
```

Listing 5.3: Prepositions in Moving Ego Constructions

This list of prepositions is retrieved via analysis of the COCA corpus. The excel screenshot shows which prepositions are used with which motion verbs (Figure 5.1). An "x" in the column means that an instance of Moving Ego metaphor is found in combination with the respective preposition. The preposition "in" was removed from the list even though sentences such as (5.13) were found in the COCA corpus.

(5.13) Because we can not travel back **in** time.

VERB	**Dir. Object**	**No Dir. Object**	No Prep	to	at	through	toward(s)	into	from	in	
approach	x		x								we **approach** the **time** for trial.
arrive	x				x						It is too easy to **arrive** at the **end** of a unit or project
creep	x						x				It's **creeping** toward dinner **time**.
enter	x		x					x			high school students before they **enter** into adulthood
escape	x								x		He ends by describing his **escape** from the **time** period
exit	x		x								the **era** which some people think we are **exiting**
flee	x		x								US now are **fleeing** economic hard **times**.
get close	x			x							it **gets close** to the harvest **time**
get near	x		x								it's **getting near** dinner **time**
go	x			x		x	x	x			we **go** back in **time** to 1969
head	x			x			x	x			we **head** into the busiest **time** of year for hotels
leave	x		x								My family had never really **left** the 19th **century**
return	x			x							and **return** to a **time** of military dictatorship
step	x							x			**Step** into summer with CLOGS.
travel	x					x					If God had wanted us to travel through time
walk	x					x					she was going to **walk** **through** **time**
wander	x					x					her mind **wandered** free through **time**.
move	x			x		x	x	x			we are **moving** closer to a **time** in

Figure 5.1: Moving Ego Analysis with Examples from COCA

This is done in order to avoid wrong matches that occurred due to the idiomatic expression "in time" which is synonymous with "punctual". Examples of Moving Ego constructions are given in the far right corner. The program should manage to detect Moving Ego sentences such as "It's not just Teaneck that's been going through tough times". The relevant segment of the dependency parse of this sentence is given in example (5.14).

(5.14) ... **going through tough times**.

...
amod(time-12, tough-11)
prep_through(motion_mE:going-9, time-12)
dep(time-12, times-14)

It is the second dependency that indicates that the sentence contains a Moving Ego expression. The rule for detecting the primary dependency of the Moving Ego metaphor is shown in Listing 5.4.

```
1 m/(prep_($prepositions)|dobj)\(motion_mET?:(.+? -- \d+), (time -- \d+)\)/
```

Listing 5.4: RegEx: Moving Ego – Primary Dependency

The danger of this approach is that some sentences which contain a non-temporal direct object and a temporal adjunct are returned as well. The sentence "when I walked through the Oval Office door at a time of maximum peril in our economy" contains the the dependencies (5.15) and (5.16).

(5.15) prep_through(motion_mE:walked-7, door-12)

(5.16) prep_at(motion_mE:walked-7, time-16)

Dependency (5.16) would be matched and a non-metaphorical sentence would be mistakenly returned. Here again, a conditional rule that no other non-temporal object or adjunct could be present is added (Listing 5.5).

```
1 m/((prep_.+?|dobj)\(motion_mET?:$motionVerb-$verbIndex, (?!time))/
```

Listing 5.5: Moving Time – Conditional Rule (1)

The current pseudocode for detecing the Moving Ego primary dependency is shown in (5.17).

(5.17) *If the program matches dobj(motion verb, time) or prep_(motion verb, time):*
- If there is a dependency between the same motion verb and
another non-temporal object: Ignore sentence
- Else: Return the sentence as Moving Time metaphor!

The addition of this conditional rule is essential as many sentences contain objects and temporal adjuncts. This extended code ensures that that improvement is made to the same degree as in the section of detecting Moving Time constructions.

The second and last challenge was to process "going to" constructions correctly. It was necessary to filter out this very frequent construction, without taking the risk of overseeing sentences such as "going back to your childhood". The phrase "going to be a problem later, at any point." with the "going to" + infinitive construction is parsed in the following manner (5.18).

(5.18) "**going to be a program later, at any point**."

...
aux(be-19, to-18)
xcomp(motion_mET:going-17, be-19)
(...)
prep_at(motion_mET:going-17, time-26)
appos(time-26, point-28)

Whenever the program detects a dependency that triggers off the Moving Ego metaphor, it rescans the sentence and searches after the dependency combination of "aux(infinitive, to)" and "xcomp(going, infinitive)". If the "going" verb and the infinitive verb are co-indexed, then this case is ignored. The regular expression for detecting "going to" constructions is given in Listing 5.6.

```
1 m/aux\((.+?-\d+), to-\d+\)\s(.+?\s)?xcomp\(motion_mET?:going-$verbIndex, .+?-\d+\)/
```

Listing 5.6: Moving Time – Conditional Rule (2)

With this second conditional rule all irrelevant "going to" cases could be filtered out successfully. Sentence which have "go" as the infinitive, e.g. "We're not going to **go** back to those days.", are successfully analyzed as instances of the Moving Ego metaphor (5.19).

(5.19) "**We're not going to go back to those days**."
…
motion_mET:go-6, to-5)
xcomp(motion_mET:going-4, motion_mET:go-6)
(…)
prep_to(motion_mET:go-6, time-10)
appos(time-10, days-12)

So in total, the Moving Ego Detector needs two conditional rules to remove unwanted sentences. The final pseudo-code is (5.20)

(5.20) *If the program matches dobj(motion verb, time) or prep_(motion verb, time):*
- If the motion verb is part of the "going to" construction: Ignore Sentence
- Else: If there is a dependency between the same motion verb and another non-temporal object: Ignore sentence
- Else: Return sentence!
- Else: If there is a dependency between the same motion verb and another non-temporal object: Ignore sentence
- Else: Return sentence!

Examples (5.21 – 5.24) illustrate the output of Moving Ego expressions containing the primary dependency.

(5.21) It's not New Jersey that's been [going] through tough [times: TIME]. - MOVING EGO

(5.22) We're not going to [go] back to those [days: TIME]. - MOVING EGO

(5.23) And we spoke then about how, after [years: TIME] of failed policies here in Washington, after [decades: TIME] of putting off the toughest challenges, we had finally [reached] a tipping [point: TIME] a [point: TIME] where the fundamental promise of America was at risk. - MOVING EGO

(5.24) That can't be the kind of leadership that we need [going] into the 21st [century: TIME]. - MOVING EGO

5.1.3 Time Orientation Dependencies

Two primary dependencies have to be captured for the Time Orientation metaphor. In the first dependency, the ego plays an active role by having time units in his sight, as illustrated in examples (2.5) and (2.6). The dependency structure of this pattern is very similar to the one of the Moving Ego metaphor. In this case the temporal unit must be the object of a verb of sight, instead of a motion verb, which is expressed with one of these four dependencies:

- **prep_on**(verb of sight, time)
- **prep_at**(verb of sight, time)

- **prep_to**(verb of sight, time)
- **dobj**(verb of sight, time)

The prepositions “on”, “at” and “to” are saved in the variable “$prep_Sight” (Listing 5.7).

```
my $prep\_Sight = ``on|at|to";
```

Listing 5.7: Prepositions in Time Orientation Constructions

The primary dependencies of the Time Orientation metaphor with verbs of sight have structures such as (5.25) and (5.26).

(5.25) dobj(sight_TO:face-4, time-8)

(5.26) prep_at(sight_TO:looking-3, time-7)

These types of dependencies are captured with the regular expression in Listing 5.8.

```
m/(prep_($prep_Sight)|dobj)\(sight_TO:(.+?)-(\d+), (time)-(\d+)\)/
```

Listing 5.8: RegEx: Time Orientation – Primary Dependency (Sight)

In order to avoid erroneous output due to multiple object or adjunct constituents as in (5.27) a conditional rule that no non-temporal object or adjunct can be linked to the verb of sight is needed again.

(5.27) “**Did you look at yourself that day?**”
…
root(ROOT-0, sight_TO:look-3)
prep_at(sight_TO:look-3, yourself-5)
det(time-7, that-6)
dobj(sight_TO:look-3, time-7)

The additional conditional rule is very similar to the first one in the Moving Ego section (Listing 5.5). The important difference lies in the replacement of “motion_mET?” with “sight_TO”, as illustrated in Listing 5.9.

```
m/((prep_.+?|dobj)\(motion_mET?:$motionVerb-$verbIndex, (?!time))/
```

Listing 5.9: Time Orientation (Sight) – Conditional Rule

The final pseudocode for capturing the grammatical relations between a verb of sight and a time unit is described in (5.28).

(5.28) *If the program matches dobj(verb of sight, time) or prep_(verb of sight, time):*
- If there is a dependency between the same verb of sight and
another non-temporal object: Ignore sentence
- Else: Return the sentence as Time Orientation metaphor!

A few examples of the first primary dependency with verbs of sight are given in (5.29 – 5.32).

(5.29) One [year: TIME] ago, we were [looking] at the possible [end: TIME] of General Motors. - TIME ORIENTATION

(5.30) And I have wondered, if Mariah is lucky enough to live as long as Marguerite Lewis 105 [years: TIME] old if she someday has a chance to [look] back on the 21st [century: TIME], what will she see. - TIME ORIENTATION

(5.31) If they want to go tell Theresa that once again you could [face] a [lifetime: TIME] of debt if you lose a family member, they can run on that platform. - TIME ORIENTATION

(5.32) Along the way, I'm sure you [faced] a few [moments: TIME] when you asked yourself: What am I doing here. - TIME ORIENTATION

In the second primary dependency, the location of a time unit is described with respect to the alignment and position of the ego, e.g. "behind", "ahead" and "in front of". The ego, which serves as the point of orientation, is either overtly mentioned, e.g. with the object pronoun "us" (5.33 – 5.34), or omitted in the sentence (5.35 – 5.36).

(5.33) We hope that the worst is behind us.

(5.34) … confident that our best days are still ahead of us.

(5.35) There is a great day ahead.

(5.36) … brighter days are still ahead.

Listing 5.10 gives an overview of all the components that are needed in order to capture these sentences. The first variable "$triggers_TO" includes words and phrases that describe the location of a temporal unit with respect to the ego. The other two variables contain all object pronouns and all word forms of the copula "be".

```
my $triggers_TO = ''ahead|behind|in_front_of|ahead_of";
my $obj_pronouns = ''me|you|him|her|it|us|them";
my $be = ''are|is|was|were|been";
```

Listing 5.10: Components of Ahead/Behind Dependencies

Ahead and behind constructions with object pronouns are captured with the first regular expression in Listing 5.11 and those without object pronouns are retrieved with two regular expressions given in Listing 5.12.

```
m/prep_($triggers_TO)\(($be)-\d+, ($obj_pronouns)-\d+\)/
```

Listing 5.11: RegEx: Time Orientation – Primary Dependency (with Object Pronoun)

```
m/nsubj\(($be)-(\d+), (time|event)-\d+\)(\s|\S)+advmod\(($be)-(\d+), ($triggers_TO)-\
    d+\)/
m/npadvmod\(($triggers_TO)-\d+, (time|event)-\d+\)/
```

Listing 5.12: RegEx: Time Orientation – Primary Dependency (without Object Pronoun)

Samples of the output are shown in (5.37 – 5.40).

(5.37) And in [times: TIME] like these, questions have always arisen about whether or not America's best [days: TIME] are [behind us]. - TIME ORIENTATION

(5.38) And not all the difficult [days: TIME] are [behind us]. - TIME ORIENTATION

(5.39) And there are going to be some hard [days: TIME] [ahead]. - TIME ORIENTATION

(5.40) In [days: TIME] of hardship, they renew our hope that brighter [days: TIME] are still [ahead]. - TIME ORIENTATION

The current version of the CTM Detector contains only rules for capturing primary dependencies. It can be considered as a preliminary version of the CTM Detector as an updated version of this program, scoring higher performance rates, will be introduced in the next sections. From now on, the current version will be referred to as the "beta version" and the updated program including all the extensions and modifications will be called the "final version" of the CTM Detector.

	Total
Time. Orient.	123
Mov. Time	186
Mov. Ego	178
	487

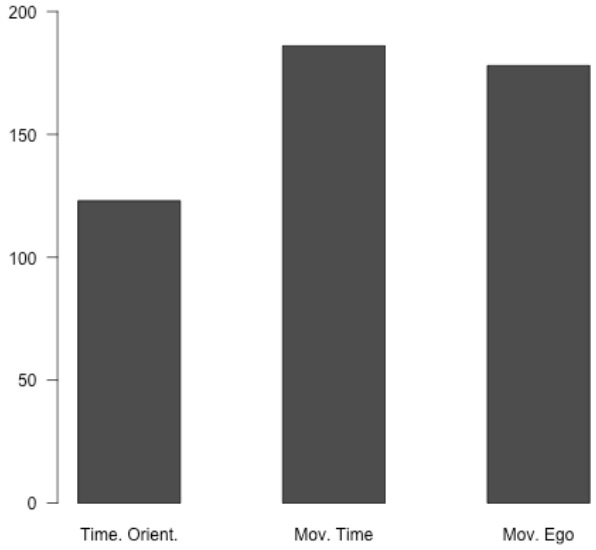


Figure 5.2: Frequencies of the Conceptual Time Metaphors in the Obama Gold Standard Database

5.2 Gold Standard Database

In order to evaluate the performance of the beta version, its output needed to be compared to a gold standard. The "Gold Standard Database"[2] is a manually compiled collection of conceptual time metaphors that were found in all of the Obama speeches. The speeches were analyzed by hand and conceptual time metaphors occuring with time or event nouns were saved in one of the three spreadsheet files for the Time Orientation, Moving Ego or Moving Time metaphor. The compilation work was supported by the Time/Event Tagger that tagged and highlighted temporal nouns. The frequencies of the conceptual time metaphors are shown in Figure 5.2 and will be referred to as the "Gold Standard Values". 487 instances of conceptual time metaphors were found in total, with Moving Time and Moving Ego metaphors representing the most common types.

The performance of the beta version was tested by running the CTM Detector on the Obama corpus and comparing its output with the Gold Standard Database. Figure 5.3 presents the precision rates for detecting each type of conceptual time metaphor. In total 240 sentences were retrieved and 214 instances were also found in the Gold Standard Database. In statistical terms, the beta version achieved a precision rate of 89.17%. The regular expressions for detecting Time Orientation metaphors proved to be the most reliable ones (95.76%), followed by rules of the Moving Time (86.27%) and Moving Ego (80.28%) metaphors.

However, a comparison between the frequencies of true positives of the beta version and the Gold Standard Values reveals that only 43.85% of all conceptual time metaphors were correctly identified, as illustrated in Figure 5.4. The juxtapositioning of the results show that the rules for detecting primary dependencies of the Time Orientation Metaphor are also reliable with respect to recall. 91.87% of all detected Time Orientation metaphors contained either dependencies with verbs of sight or temporal descriptions with e.g. "ahead" and "behind". It was a great surprise that the majority of Moving Ego and Moving Time metaphors could not be detected with the primary dependency rules which were thought to represent stereotypical structures of motion time metaphors. Only 23.66% (Moving Time) and 32.02% (Moving Ego) of relevant metaphors were retrieved with the beta version.

As a consequence, two further steps were needed in order to improve the CTM Detector. All the sentences

[2]FILES: "**GS_TimeOrientation.xls**", "**GS_MovingTime.xls**", "**GS_MovingEgo.xls**" — LOCATION: "**./GoldStandard_DB**"

	True.Pos.	Marked.Pos.	%
TO	113	118	95.76
MT	44	51	86.27
ME	57	71	80.28
	214	240	89.17

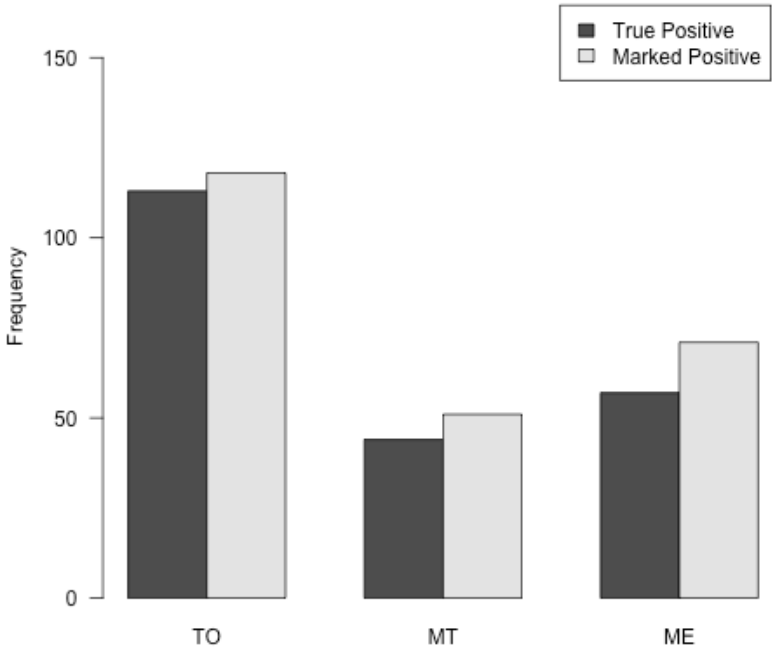


Figure 5.3: Precision – Only Primary Dependency Rules (Beta Ver.)

that were not matched were collected manually and grouped together whenever they contained similar structures. The second step involved the creation of regular expressions for detecting as many instances of these groups as possible. The next section will present the measurements that were taken in order improve the CTM Detector.

5.3 Extensions and Modifications

5.3.1 Moving Time: Reduced Relative Clause

The first extension rule captures sentences containing reduced relative clauses as in (5.41 – 5.44).

(5.41) And I'm going to keep on needing him in the Senate in the years **to come**.

(5.42) And that will create thousands of jobs across the country thousands of jobs across the country, not just this year, not just next year, but for decades **to come**.

(5.43) That's the strategy that we're going to be pursuing in the months and years **to come**.

(5.44) That's what I'm going to ask you to continue in the weeks and months and years **to come**, as fellow travelers in this effort for us to perfect our union.

These relative clauses are not initialized with relative pronouns or complementizers, and are normally attached right behind the temporal noun. The Stanford Parser labels the grammatical relation between the time/event noun and the motion verb as *auxiliary* (= "aux"). These sentences are retrieved with the regular expression in Listing 5.13.

	True.Pos.	Gold.St.	%
TO	113	123	91.87
MT	44	186	23.66
ME	57	178	32.02
	214	487	43.94

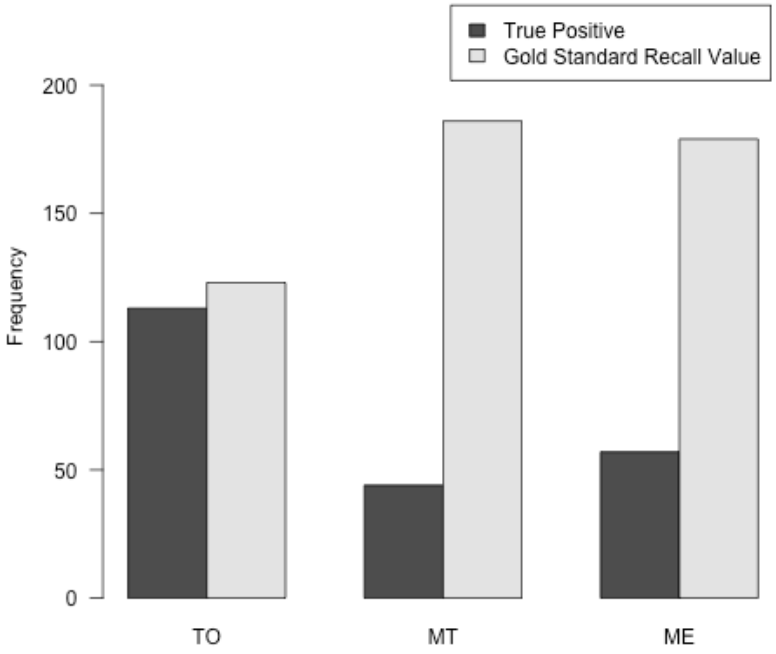


Figure 5.4: Recall – Only Primary Dependency Rules (Beta Ver.) Compared to Gold Standard

```
m/appos\((time|event)-\d+, .+?-(\d+)\)\naux\((motion_mE?T:(.+?)-\d+), to-(\d+)\)/
```

Listing 5.13: RegEx: Moving Time – Reduced Relative Clause

The examples (5.41 – 5.44) are returned as in (5.45 – 5.48).

(5.45) And I'm going to keep on needing him in the Senate in the [years: TIME] to [come]. - MOVING TIME

(5.46) And that will create thousands of jobs across the country thousands of jobs across the country, not just this [year: TIME], not just next [year: TIME], but for [decades: TIME] to [come]. - MOVING TIME

(5.47) That's the strategy that we're going to be pursuing in the [months: TIME] and [years: TIME] to [come]. - MOVING TIME

(5.48) That's what I'm going to ask you to continue in the [weeks: TIME] and [months: TIME] and [years: TIME] to [come], as fellow travelers in this effort for us to perfect our union. - MOVING TIME

5.3.2 Moving Time: Adjectival Participle

The second Moving Time extraction rule contains a regular expression that captures deverbal adjectives as in (5.49 – 5.52).

(5.49) And they know that the actions we honor today were not a **passing** moment of courage.

(5.50) Now, it is our fervent hope that in the **coming** weeks, Chrysler will find a viable partner and GM will develop a business plan that will put it on a path to profitability without endless support from American taxpayer.

(5.51) He told me he'd joined the Marines and was heading to Iraq the **following** week.

(5.52) In the **coming** days and weeks, I'll be launching other aspects of the plan.

These examples prove that the Moving Time metaphor is not only expressed with motion verbs but also with adjectival participles. For this reason, the dependency between motion adjectives and time/event nouns is either labeled as *adjectival modifier* (= "amod") or as *participal modifier* (= "partmod"). These modifier dependencies are matched by the regular expression in Listing 5.14.

```
1 m/(amod|partmod)\((time|event)-\d+, motion_mE?T:(.+?)-\d+\)/
```

Listing 5.14: RegEx: Moving Time – Adjectival Participle

Metaphorical utterances are returned as illustrated in (5.53 – 5.56).

(5.53) And they know that the actions we honor today were not a [passing] [moment: TIME] of courage. - MOVING TIME

(5.54) Now, it is our fervent hope that in the [coming] [weeks: TIME], Chrysler will find a viable partner and GM will develop a business plan that will put it on a path to profitability without endless support from American taxpayer. - MOVING TIME

(5.55) He told me he'd joined the Marines and was heading to Iraq the [following] [week: TIME]. - MOVING TIME

(5.56) In the [coming] [days: TIME] and [weeks: TIME], I'll be launching other aspects of the plan. - MOVING TIME

5.3.3 Moving Time: Go By

The last additional rule for detecting Moving Time metaphors is the most specific one. A closer look at the Obama speeches showed that the motion verb "go" can, in fact, occur in Moving Time metaphors if it co-ocurrs with "by" as a phrasal verb. This is illustrated with the examples (5.57) and (5.58).

(5.57) But each day **went by** without any orders to halt construction so they kept on working and they kept on building.

(5.58) The costs of health care are not going to come down overnight just because legislation passed, and in an ever-changing industry like health care, we're going to continuously need to apply more cost-cutting measures as the years **go by**.

The regular expression for detecting this type of Moving Time metaphor is given in Listing 5.15.

```
1 m/nsubj\((motion_mE:(go|went|gone|going|goes))-\d+, (time|event)-\d+\)(\s|\S)+(advmod
      |acomp|prep)\((motion_mE:(go|went|gone|going|goes))-\d+, ($adverbs)-\d+\)/
```

Listing 5.15: RegEx: Moving Time – Go By

The field "motion_mE" needs to be mentioned explicitly in the regular expression in order to overwrite the standard lexical entry that all word forms of "go" can only occur in Moving Ego metaphors. The variable "$adverbs" contains, inter alia, the adverb "by". Examples (5.59) and (5.60) show that both components of the phrasal verbs are tagged.

(5.59) But each [day: TIME] [went by] without any orders to halt construction so they kept on working and they kept on building. - MOVING TIME

(5.60) The costs of health care are not going to come down overnight just because legislation passed, and in an ever-changing industry like health care, we're going to continuously need to apply more cost-cutting measures as the [years: TIME] [go by]. - MOVING TIME

5.3.4 Moving Ego: Non-Motion Verb Get

The Obama corpus showed that the ego can "get" to a "point" within the Moving Ego time landscape, as illustrated in (5.61) and (5.63).

(5.61) Now, some may have thought it would take months to **get** to this point.

(5.62) How do we **get** ourselves to the point where more people are working, and more people are spending, and you start seeing a virtuous cycle and the recovery starts to feed on itself.

(5.63) I'm telling you, I'm **getting** to the point where I'm not going to get applause.

This case is interesting as "get" is conventionally not associated with motion and that a word such as "point" was tagged as an expression of time by the Time/Event Tagger ("[point: TIME]"). One could argue whether "point" is really referred to a discrete moment within the time landscape or whether it refers to a "point" on a different dimension. Support however is found that they are expressions of the Moving Ego metaphor as sentences such as (5.64) are also found in the Obama speeches.

(5.64) friends who have helped get me to this **point in time**.

This quote by Obama makes explicit that reference is made to a specific point in the dimension of time. A regular expression was created for capturing Moving Ego constructions that contain the non-motion verb "get" (Listing 5.16).

```
1 m/prep_to\((\$get)-\d+,time-\d+\)/
```

Listing 5.16: RegEx: Moving Ego – Non-Motion Verb Get

The examples given in this section so far are returned by the CTM Detector as illustrated in (5.65) – (5.68)

(5.65) friends who have helped [get] me to this [point in time: TIME]. - MOVING EGO

(5.66) Now, some may have thought it would take [months: TIME] to [get] to this [point: TIME]. - MOVING EGO

(5.67) How do we [get] ourselves to the [point: TIME] where more people are working, and more people are spending, and you start seeing a virtuous cycle and the recovery starts to feed on itself. - MOVING EGO

(5.68) I'm telling you, I'm [getting] to the [point: TIME] where I'm not going to get applause. - MOVING EGO

5.3.5 Moving Ego: Through/Throughout

Normally one might assume that Moving Ego metaphors are formed with motion verbs or with exceptions such as "get". A closer analysis of the other missing Moving Ego sentences however revealed that the ego can actually move in time while doing something else. This is illustrated in examples (5.69 – 5.78).

(5.69) **navigate**
- In so many ways , each of you today have shown that ingenuity as you 've successfully navigated your companies through an extraordinarily difficult time the toughest time that we 've seen since the Great Depression.

(5.70) **help**
- We gave relief to states to help them through these tough times.
- And I believe they will be a lifeline to help viable small businesses through these difficult times.

(5.71) **work**
- So I want to congratulate you for having worked through a very difficult year.
- The United States is still working through some of our own darker periods in our history.
- Each country must work through its past.
- And the best way forward for the Turkish and Armenian people is a process that works through the past in a way that is honest , open and constructive.

(5.72) **serve**
- As governors , I know you feel the same responsibility to see the people we serve through difficult times.

(5.73) **lead**
- He 's led Mary Queen of Vietnam Church in Louisiana through some pretty hard days.
- Now , like his father and his grandfather before him , the Prime Minister is leading Greece through challenging times.
- Congratulations to the owner , Tom Benson , who has led this team through times that would test anybody ; and General Manager Mickey Loomis , for building this extraordinary championship squad.

(5.74) **plow**
- They plowed through the regular season , they won every game in the AFC North , they took down the Chargers and the Ravens in the playoffs.

(5.75) **carry**
- It 's the belief that has brought millions of people to our shores , and carried us through even the toughest economic times.
- So plenty of cities carry their sports teams through a tough season.
- It 's a rare thing when a sports team carries a city through tough times.

(5.76) **make a way**
- Together , we shall make a way through winter , and we 're going to welcome the spring.

(5.77) **make it**
- extending unemployment insurance and COBRA to help folks make it through some really tough times

(5.78) **live**
- We 're living through extraordinary times.

This list of examples reveals the wide range of verbs that can be used to express the Moving Ego metaphor. It would not have made sense to add each of these verbs into VerbLex as they do not denote locomotion. One of the characteristics that all of these examples have in common is that their temporal nouns bear the Duration Sense. The preposition "through" indicates that the ego is actually doing something in the entire time while going towards the end of a temporal room. Sentences such as (5.69 – 5.78) can be successfully captured if the assumption is raised that the combination between "through" and a temporal noun always represents instances of the Moving Ego metaphor. This idea is also based on the definition given by the New Oxford American Dictionary in (5.79).

(5.79) **through**
"continuing in time toward completion of (a process or period): [as prep.] : he showed up halfway through the second act" (New Oxford American Dictionary, 2001)

From the perspective of the Contemporary Metaphor Theory, the definition says that the ego *moves along toward the offset of a time period*. In the end, the decision was made to create an extra regular expression for detecting dependencies between the preposition "through" and temporal nouns. Modifications in the previously mentioned lexica and regular expressions were required. The list of prepositions in the beta version (Listing 5.3) needed to be divided into two groups, as illustrated in Listing 5.17. The first variable "$prep_MomentSense" contains all directional prepositions that trigger off the Moment Sense and the second variable "$prep_DurationSense" contains "through"[3].

```
my $prep_MomentSense = "to|at|towards?|from|into";
my $prep_DurationSense = ''through'';
```

Listing 5.17: Prepositions in Moving Ego Constructions (FINAL Version)

Consequently, the regular expression for detecting the primary dependency of the Moving Ego metaphor (Listing 5.4) needed to be adjusted. The modified regex is now limited to capturing instances of the Moving Ego metaphor containing the Moment Sense (Listing 5.18).

```
m/(prep_($prep_MomentSense)|dobj)\(motion_mET?:(.+?)-(\d+), (time)-(\d+)\)/
```

Listing 5.18: RegEx: Moving Ego – Primary Dependency (FINAL Version)

The new extension rule for capturing "through" constructions is presented in Listing 5.19.

```
m/prep_($prep_DurationSense)\((.+?)-\d+, time-\d+\)/
```

Listing 5.19: RegEx: Moving Ego – Through

Examples of the output of this regular expression are given in (5.80 – 5.83).

(5.80) And that's completely understandable because we are going [through] the toughest economic [times: TIME] in our living memory. - MOVING EGO

(5.81) America we are passing [through] a [time: TIME] of great trial. - MOVING EGO

(5.82) He's led Mary Queen of Vietnam Church in Louisiana [through] some pretty hard [days: TIME]. - MOVING EGO

(5.83) It's the belief that has brought millions of people to our shores, and carried us [through] even the toughest economic [times: TIME]. - MOVING EGO

5.3.6 Time Orientation: Lie Ahead/Behind

The evaluation of the beta version of the CTM Detector showed that the extraction of primary dependencies of the Time Orientation metaphor were very reliable (Precision: 95.76% – Recall: 91.87%). The only minor upgrade that raised the performance of detecting Time Orientation metaphors was the addition of a regular expression that captures dependencies between the verb "lie" and temporal nouns. Sentences such as (5.84 – 5.87) could not be captured with the primary dependency rules.

(5.84) ...and that here in America, our best days lie ahead.

(5.85) Great days lie ahead for this nation.

(5.86) And I am confident that better days lie ahead.

(5.87) I'm convinced that the choice between profit and progress is a false one and that the golden days of journalism still lie ahead.

[3]The preposition "into" is not part of to the Duration Sense group as focus is only set on motion past the *onset* of a period. (Defined as Moment Sense in Section 2.3)

The regular expression for capturing these sentences is presented in Listing 5.20.

```
m/nsubj\(($lie)-(\d+), (event|time)-\d+\)(\s|\S)+advmod\(($lie)-(\d+), ($triggers_TO)
    -(\d+)\)/
```

Listing 5.20: RegEx: Time Orientatin – Lie Ahead/Behind

Examples (5.84 – 5.87) are returned as (5.88 – 5.91)

(5.88) That we can make of our lives what we will: that all things are possible for all people: and that here in America, our best [days: TIME] [lie ahead]. - TIME ORIENTATION

(5.89) Great [days: TIME] [lie ahead] for this nation. - TIME ORIENTATION

(5.90) And I am confident that better [days: TIME] [lie ahead]. - TIME ORIENTATION

(5.91) I'm convinced that the choice between profit and progress is a false one and that the golden [days: TIME] of journalism still [lie ahead]. - TIME ORIENTATION

In conclusion, the final version of the CTM Detector contains ten, instead of four extraction rules. The next chapter will present statistics about the performance of each rule and outline the different types of errors that were made.

Chapter 6

Results

This chapter will present statistics about the performance of the final version of the CTM detector and describe the effects of the extension rules and modifications (described in Section 5.3). A manual evaluation of the output of the final version revealed that the overall recall score increased up to 88.50%. The recall values of the beta, final and the Gold Standard version are given in Table 6.1 and they are visually presented in Figure 6.1.

	True.Pos (BETA)	%	True.Pos (FINAL)	%	Gold Standard
Time Orientation	113	91.87	117	95.12	123
Moving Time	44	23.66	161	86.56	186
Moving Ego	57	32.02	153	85.96	178
	214	43.94	431	88.50	487

Table 6.1: Recall – Beta, Final and Gold Standard Version

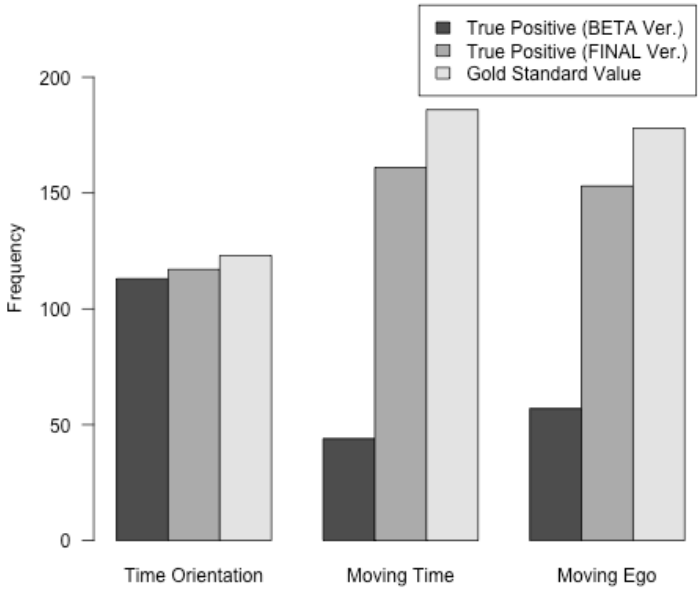


Figure 6.1: Recall – Beta, Final and Gold Standard Version

The final version of the CTM detector outperformed the beta version by more than 200% with respect to recall, as the number of true positive instances rose from 214 to 431. This drastic improvement in performance is attributed to the extensions and modifications that were mainly added for capturing further Moving

Time and Moving Ego instances. The recall values of Moving Time and Moving Ego metaphors increased by 365.91% and 268.42% respectively. This means that the number of metaphorical expressions captured by the extension rules was actually far higher than the number of marked positives by the primary dependency rules.

The overall precision rate of the final version and the precision rates of detecting each type of conceptual time metaphor are given in Figure 6.2.

	True.Pos.	Marked.Pos.	%
TO	117	122	95.90
MT	161	184	87.50
ME	153	177	86.44
	431	483	89.23

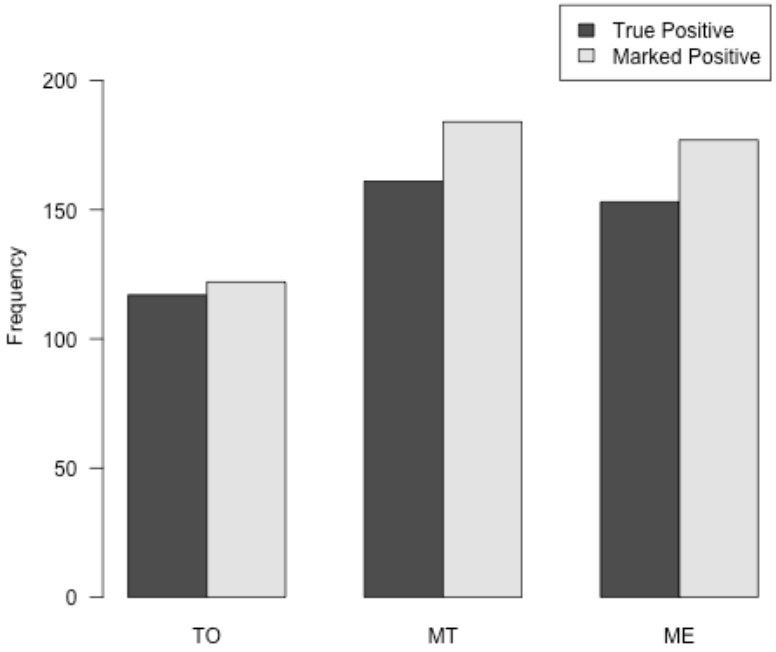


Figure 6.2: Precision – Primary Dependency Rules with Extension Rules (FINAL Ver.)

The overall precision rate lies at 89.23% which is very close to the 89.17% scored by the beta version (Figure 5.3). The rules for detecting the Time Orientation metaphor performed best again, followed by the rules for Moving Time and Moving Ego metaphors. These numbers indicate that the extension rules performed well in retrieving correct instances without raising the number of false positives too much. In order to get a better understanding of how the statistics came about, the performance of every single rule will be described in the next sections.

6.1 Time Orientation Rules

Precision

It is no surprise that only very little improvement was made in detecting Time Orientation metaphors as the primary dependencies rules proved to be highly accurate in the beta version already. Figure 6.3 shows that 84 of the 117 true positive instances of the Time Orientation rules contained the second primary dependency with adverbs, e.g. “ahead” or “behind” (abbr. “PD:AHEAD”). The first primary dependency rule with the sight verbs “look” and “face” captured 29 sentences correctly and the extension rule that captured dependencies between the verb “lie” and a temporal noun retrieved four more correct instances without raising the quantity of false positives.

	True.Pos.	Marked.Pos.	%
PD:SIGHT	29	33	87.88
PD:AHEAD	84	85	98.82
EXT:LIE	4	4	100.00
	117	122	

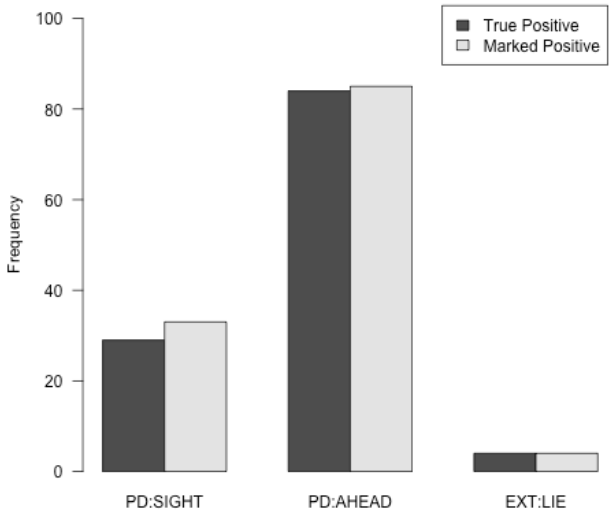


Figure 6.3: Precision – Performance of Time Orientation Rules

Type I Errors

One of the five false positives was caused due to lexical issues. The Time/Event lexicon contained and tagged nouns such as "watch" (6.1).

(6.1) …she'd be [looking] at her [watch: TIME] and thinking, … - TIME ORIENTATION

In sentence (6.1), the agent "she" is literally having a look at the time by using her "watch". The noun "watch" is retrieved from WordNet as a temporal noun as this term can refer to a period during which a person, e.g. guard or ship's crew member, is on duty.

The other four false positives were analyzed incorrectly due to erroneous analyses of the Stanford Parser. Examples (6.2) and (6.3) reveal that the Stanford Parser has issues analyzing sentences containing syntactic movements such as fronting. The direct object dependency was not created between "face" and "setbacks"/"choice", but with the temporal adjuncts "year" and "fall". An object dependency is mistakenly created between "point" and the imperative "look" in example (6.4). The last example (6.5) is actually a correct instance of the Time Orientation metaphor found in the Gold Standard database. It is however classified as a false positive as it is returned by accident due to a dependency between "behind us" and "moments".

(6.2) But I wake up every [day: TIME] knowing that they are nothing compared to the setbacks that families all across this country have [faced] this [year: TIME]. - TIME ORIENTATION

(6.3) That's the choice we [face] this [fall: TIME]. - TIME ORIENTATION

(6.4) And at some [point: TIME] you had to make a decision and then you've got to tell your employees, [look], I know it's right to cover you, … - TIME ORIENTATION

(6.5) We hope that the worst is [behind us], but it's at [moments: TIME] like this where leadership is tested. - TEMPORAL SEQUENCE

Type II Errors

The Time Orientation rules did not manage to detect six instances. One of the false negatives is the previously mentioned example (6.5). This example reveals that the CTM Detector has its weaknesses in detecting nominalized adjectives such as "the worst" as a potential component of the Time Orientation metaphor. Sentence (6.6) is not retrieved, as the verb "see" is not included in VerbLex. The motivation behind not having "see" in the lexicon was described in *Motion/Sight Tagger* (Section 4.4).

(6.6) You will see a [time: TIME] in which we as a nation finally recognize …

The four other false negatives were caused by erroneous outputs of the Stanford Parser. Two very specific cases will be presented. Example (6.7) could have been successfully captured if the Stanford Parser produced a dependency between "years" and "ahead", but a dependency is only found between "tests" and "ahead".

(6.7) To all of you from near and far , and over all the [years: TIME] and tests [ahead] , ….

The Stanford Parser struggled also with multi-token specifiers such as "one of those" as illustrated in (6.8).

(6.8) **[face] one of those [moments: TIME]**

dobj(sight_TO:face-5, one-6)
det(time-9, those-8)
prep_of(one-6, time-9)

The direct object dependency is not created between the sight verb and "moments", as the word "one" is mistakenly analyzed as the head of the noun phrase and "of those moments" is described as its modifier.

6.2 Moving Time Rules

Precision

Figure 6.4 reveals how each of the Moving Time rules performed. The results show that the extension rules for detecting reduced relative clauses ("EXT:RRC") and adjectival participles ("EXT:ADJ") were the main reasons for the drastic improvement in performance. The RRC rule retrieved most of the Moving Time instances (total: 75) and was one of the most accurate one (91.46%). The extension rule for detecting "go by" constructions played only a minor role in contributing to the performance of the CTM detector. However, every rule that added to the numbers of positive matches without raising the numbers of false positives was kept.

Type I Errors

23 Type I errors were found in the output of the Moving Time rules. Nine of them were caused by lexical challenges such as lexical ambiguity or idiomaticity. The extension rule for adjectival participles returns six instances which contain the idiomatic expression "in the near term". Examples (6.9) and (6.10) contain this phrase.

(6.9) And the first round of funding will focus on projects that can create jobs and benefits in the [near] [term: TIME]. - MOVING TIME

(6.10) What I'm interested in is taking action right now to help businesses create jobs right now, in the [near] [term: TIME]. - MOVING TIME

It is interesting that the CTM Detector returned the phrase "in the near term" as it resembles structures of a conceptual time metaphor at first sight. One could even suggest that it is based on a conceptual time metaphor. In this study however, these sentences are not regarded as instances of the Moving Time metaphor as this phrase is synonymous to "soon" and no motion is involved.

Example (6.11) is returned due to a dependency between "generation" and the verb "come". It is not an expression of the Moving Time metaphor as the idiomatic expression "to come of age" carries the meaning "to reach maturity".

	True.Pos.	Marked.Pos.	%
PD:MOTION	44	51	86.27
EXT:RRC	75	82	91.46
EXT:ADJ	40	49	81.63
EXT:Go By	2	2	100.00
	161	184	

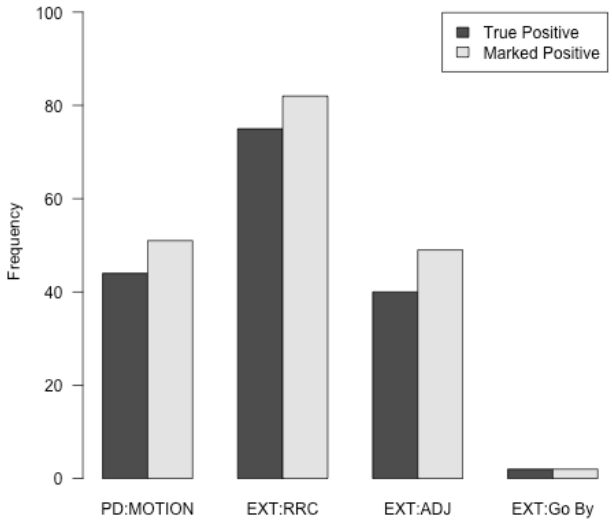


Figure 6.4: Precision – Performance of Moving Time Rules

(6.11) They're a [generation: TIME] that [came] of [age: TIME] amidst the horrors of 9 and Katrina ... - MOVING TIME

The phrase "to run a program" is captured as an instance of Moving Time metaphor even though it is not compatible with the structures and models of any conceptual time metaphor. In contexts such as (6.12) and (6.13) "run" is not to be understood as a motion verb as it is paraphrased as "take place".

(6.12) ...who makes this [program: EVENT] [run] from the assistants to the trainers, ...- MOVING TIME

(6.13) ...the magnitude of this thing, this [program: EVENT] has [run] cleanly, smoothly, transparently. - MOVING TIME

Another special case of lexical ambiguity was found in example (6.14) where "moving terms" actually means "emotional words".

(6.14) Tony Coelho, who was instrumental on this issue, spoke in just incredibly [moving] [terms: TIME] ... - MOVING TIME

Six Type I errors that were caused by RRC rule could have been avoided if it was extended with the conditional rule that no other non-temporal subject or object can stand in a grammatical relation with the same motion verb. The first two examples (6.15) and (6.16) would have been blocked as they contain object dependencies between "pass" and "health care".

(6.15) Now is the [time: TIME] to [pass] health care. - MOVING TIME

(6.16) It is [time: TIME] to [pass] health care reform for America, ... - MOVING TIME

Despite the fact that the precision rate would have gone up to almost 100% the decision was made to not keep the conditional rule as seven positive instances of reduced relative clauses would not have been captured. So in this case, the priority was set to attain as many positive as possible at the expense of lower precision rates.

The other eight false positive instances were caused by various erroneous outputs of the Stanford parser, e.g. examples (6.17) and (6.18). A subject dependency should have actually been built between "Clinton" and "reaching" in (6.17), and between "walls" and "come" in (6.18).

(6.17) It's why you've seen Secretary Clinton in so many countries at town halls, on local [television programs: EVENT], [reaching] out to citizens ... - MOVING TIME

(6.18) But more than any other nation, the United States of America has underwritten global security for over six [decades: TIME] a [time: TIME] that, for all its problems, has seen walls [come] down ...- MOVING TIME

Type II Errors

The conditional rule that prohibits other non-temporal subjects and objects in the primary dependency rule helped to reduce the number of false positives but also brought forth the four false negatives in (6.19 – 6.22).

(6.19) ...**We ca n't let this moment pass us by**.
dobj(motion_mET:pass-15, us-16)

(6.20) **But in the words of Scripture, the time has come to set aside childish things**.
prep_in(motion_mT:come-14, words-4)

(6.21) ...**because the time has come for the world to move in a new direction**.
prep_for(motion_mT:come-39, world-42)

(6.22) ...**and the time has come for those walls to come down**.
prep_for(motion_mT:come-22, walls-25)

These examples demonstrate that the conditional rule works for most cases of the Moving Time metaphor but not always. Example (6.19) is a proof that the ego can actually refer to himself with the object pronoun "us", and the other three instances reveal that references can be made to where, "in the words" (6.20), and for whom, "world" and "walls" (6.21 – 6.22), the time has come.

Sentences (6.23 – 6.26) could not be captured due to missing Moving Time verb entries, such as "hasten" and "go".

(6.23) To hasten the day when our troops will leave.

(6.24) I'll bet they had the same feelings that you do you're a little sad to see the summer go, but you're also excited about the possibilities of a new year.

(6.25) And rarely has a day gone by that ...

(6.26) Long gone are the days when a ...

The first sentence (6.23) is an interesting realization of the Moving Time metaphor as the timing of "day" is accelerated and moved forward, closer to the ego. Examples (6.24 – 6.26) reveal that in some specific contexts Moving Time metaphors can come with the motion verb "go" without any adverbs, e.g. "away" or "by". This might have something to do with the combination of the Time Orientation metaphor "see the summer" and the Moving Time metaphor "summer go" in example (6.24), or with the adjectival participle form in (6.25 – 6.26).

The other 19 Type II errors were not retrieved due to different types of errors in the Stanford Parser output. Example (6.27) is a further proof that the Stanford Parser has problems handling syntactic movements. The adverb "then" is returned as the subject and "day" as the direct object in sentence (6.27). In (6.28),"yet" is analyzed as a verb even though it is an adverb.

(6.27) **...then out of this ordeal will come a better day and a brighter future** ...
nsubj(motion_mT:come-31, then-25)
...
dobj(motion_mET:come-31, time-34)

(6.28) **All that you've achieved, I believe that the CIA's best days are still yet to come.**
nsubj(yet-20, time-14)

The four sentences (6.29 – 6.32) containing the phrasal verb "go by" could not be captured. The examples show that "by" is wrongly parsed as a prepositional clausal modifier between e.g. "goes" and "think". In this reading, "by" is interpreted as "by means of". It would have been possible to retrieve all four sentences by matching the incorrect dependency parses but the decision was made against this option as the CTM Detector is meant retrieve correctly parsed sentences only.

(6.29) **And I want you to know that not a day goes by when I do n't think about those efforts of yours.**
prepc_by(motion_mE:goes-14, think-20)

(6.30) **...not a single day goes by that I do n't think about the obligation** ...
prepc_by(motion_mE:goes-15, think-21)

(6.31) **Not a minute has gone by that he has n't thought about how** ...
prepc_by(motion_mE:gone-8, thought-14)

(6.32) **Not a single day goes by where I do n't think about all the** ...
prepc_by(motion_mE:goes-8, think-14)

6.3 Moving Ego Rules

Precision

The precision scores of the Moving Ego rules are presented in Figure 6.5. First of all, it has to be pointed out that the scores of the primary dependency rule decreased compared to the scores provided by the beta version (Figure 5.3). The number of sentences marked positive reduced from 71 to 48 as the preposition "through" was removed from the preposition list and embedded into the extension rule which captures "through" constructions ("EXT:Through"). The creation of the EXT:Through rule did not only capture most of the Moving Ego sentences but also proved to be the most accurate one (91.95%). All 23 "through" instances that were captured in the beta version were also retrieved by the new extension rule and the list of true positives was expanded with 80 further sentences. The extension rule that matches "get to" constructions scored the lowest recall rate with 15 true positives but achieved a high precision rate (88.24%).

Type I Errors

In total 25 false positives were found. 10 of them can be ascribed to weaknesses of the Time/Event Tagger. Examples (6.33) and (6.34) include cases of lexical ambiguity. The nouns "fall" and "conclusion" are included in the Time/Event lexica as they can denote a "season of the year" and an "end of a process or event" respectively. In this case however they refer to "the act of falling" or "a decision/judgement".

(6.33) ...as the economy [goes] into free [fall: TIME] ...- MOVING EGO

(6.34) ...who [reach] different [conclusions: TIME] about the same things ...- MOVING EGO

Another weakness of the Time/Event Tagger is to detect proper names correctly, as examples (6.35) and (6.36) show. Temporal concepts such as "start" and "first" should be skipped when occurring in proper names, e.g. "New START Treaty" and "City First Bank".

(6.35) ...to reduce our nuclear arsenals [through] the New [START: TIME] Treaty ...- MOVING EGO

	True.Pos.	Marked.Pos.	%
PD:MOTION	34	48	70.83
EXT:Get To	15	17	88.24
EXT:Through	103	112	91.96
	153	177	

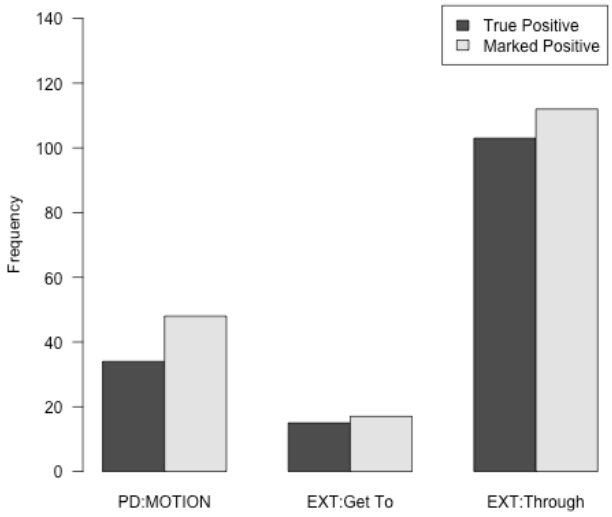


Figure 6.5: Precision – Performance of Moving Ego Rules

(6.36) Ultimately they got a loan [through] City [First: TIME] Bank, ...- MOVING EGO

Nouns such as "moon" and "bit" are found in the Time/Event lexica and cause false positives as in (6.37) and (6.38). Their temporal meanings are connected to the lunar month and a "short time", as in "in just a bit", respectively.

(6.37) We're no longer competing to achieve a singular goal like [reaching] the [Moon: TIME]. - MOVING EGO

(6.38) You could raise the tax on everybody, so everybody's payroll tax [goes] up a little [bit: TIME]. - MOVING EGO

The output of the Stanford Parser is responsible for the other 15 false positives of the Moving Ego rules. One of the common mistakes is that adjuncts are described as direct objects, as presented in examples (6.39 – 6.41). This was the main reason for the lower precision rate of the Moving Ego primary dependency rule.

(6.39) ...**his grandfather, [returning] six [decades: TIME] after he was a midshipman**, ... - MOVING EGO
dobj(motion_mE:returning-39, time-41)
appos(time-41, decades-43)

(6.40) ...**or all you midshipmen [returning] next [fall: TIME], I hereby grant you something extra an extra [weekend: TIME]**. - MOVING EGO
dobj(motion_mE:returning-39, time-41)
appos(time-41, decades-43)

(6.41) **Finally, I understand that Coach Auriemma has promised to [go] 40-0 next [season: TIME]**. - MOVING EGO
dobj(motion_mE:go-11, time-14)
appos(time-14, season-16)

In some cases, dependencies are mistakenly created between motion verbs and specifiers such as "years" in "four years of" (6.42) and "years of" (6.43). The dependents of the parses should have been "college" and "effort".

(6.42) ...**tuition relief for each child [going] to four [years: TIME] of college**. - MOVING EGO
prep_to(motion_mE:going-18, time-21)
appos(time-21, years-23)

(6.43) **He got there [through] [years: TIME] of effort**. - MOVING EGO prep_through(got-2, time-5)
appos(time-5, years-7)

Type II Errors

In total 24 true positives could not be captured by the Moving Ego rules. Six false negatives could not be captured due to missing VerbLex entries. The missing motion verbs and phrasal verbs are "come" (6.44 – 6.45), "arrive" (6.46), "emerge" (6.47), "get passed" (6.48), "take steps" (6.49) and "move closer" (6.50).

(6.44) One newspaper noted that " we have come to the hour for which we were born.

(6.45) And it captured the essence of who we are, coming out of tough times.

(6.46) For as we face down the hardships and struggles of our time, and arrive at that hour for which we were born ...

(6.47) This country is emerging from an incredibly difficult period in its history ...

(6.48) ...we want to get passed this year that says to every young person in America ...

(6.49) ...but there are critical steps we can take toward a new day.

(6.50) Together, we will have moved closer to that day when no one has to be afraid to be gay in America.

The addition of these verbs into VerbLex did not improve the performance of the Moving Ego rules. Tests have shown that the extractor performed better without deictic verbs such as "come" and "arrive" as sentences such as (6.51) with implicit deictic centers were returned as well. In most cases the implicit deictic center corresponds to the location of the addressee.

(6.51) I will come/arrive towards the end of April.

Only one instance could not be captured due to the conditional rule that no other non-temporal object or adjunct, linked to the same motion verb, could be present in the same sentence. The exception is presented in (6.52). This example demonstrates that improvements could be made by excluding prepositions such as "with" from the conditional rule.

(6.52) They joined together so that all of us could enter our golden years with some basic peace of mind.
dobj(motion_mE:enter-10, time-12)
...
prep_with(motion_mE:enter-10, peace-20)

10 sentences could not be captured as they form exceptional cases of Moving Ego expressions. Not only is it possible to "lead", "serve" or "carry" somebody through a *period of time*, as illustrated in Section 5.3.5, it is also possible to "move" (6.53), "take" (6.54) and "bring" (6.55) somebody or something to a specific *point in time*.

(6.53) **move**
- …breaking free from the politics of the past and moving America forward at this defining moment in our history.
- And it would start building the kind of infrastructure that would move America into the 21st century.
- …a foundation that will move us from an era of borrow and spend to one where we save and invest …
- …that will help us move from a period of reckless irresponsibility, a period of crisis, to one of responsibility and prosperity.

(6.54) **take**
- They want to repeal that reform, take us back to the days when insurance companies could deny you care ?

(6.55) **bring**
- Congratulations to all the parents, the cousins the aunts, the uncles all the people who helped to bring you to the point that you are here today.
- We celebrate the courage and commitment of those who brought us to this point.

Examples (6.56) and (6.57) are also exceptional cases as temporal concepts such as "months" and "years" are used as units that determine the length of the path that needs to be traveled.

(6.56) Now we've got three months to go, and so we've decided, well, we can politick for three months. ()

(6.57) I imagine there are some seniors out there who are feeling pretty good right now with just one more year to go.

Constructions like these are often used for descriptions of spatial routes. The lyrics in (6.58) are taken from the marching song replayed over and over again in the classic John Wayne movie "The Fighting Kentuckian" (1949).

(6.58) "Only 600 miles more to go" – "The Fighting Kentuckian" (1949)

The remaining seven instances could not be captured due to different types of erroneous outputs of the Stanford Parser. For example, the verb "near" is parsed as a preposition (6.59), fronting is not recognized correctly (6.60) and the according to the analysis of the Stanford Parser the Prime Minister is leading Greece *by means of* challenging times (6.61)[1].

(6.59) **Indeed, as we near Memorial Day, we pay** …
prep_near(we-4, time-6)

(6.60) **More than anyone else, the new era of service we enter in today** …
prep_in(motion_mE:enter-15, today-17)

(6.61) …**the Prime Minister is leading Greece through challenging times**.
prepc_through(leading-19, challenging-22)

[1]Very plausible interpretation but probably not what was meant …

Chapter 7

Conclusion and Future Research

The introduction stated the aim to test the statement that "the computer cannot work from a list of conceptual metaphors to identify their linguistic realizations" (Deignan, 2005). The validity of this statement was tested through the development of the CTM Detector that focuses on the retrieval of sentences containing the conceptual metaphors "Time Orientation", "Moving Time" and "Moving Ego". Sentences are preprocessed and conceptual time metaphors are extracted with a list of ten rules that capture different ways of expressing them. The final version of the CTM Detector managed to retrieve 88.50% of conceptual time metaphors in the Obama corpus. Even though not all of the instances may have been retrieved, the results reveal that the detection of conceptual metaphors is possible, at least with respect to conceptual time metaphors.

So far, the thesis is lacking an evaluation of the CTM Detector on a separate test corpus, as the statistics are only based on the Obama training corpus. Priority was given to optimizing the code of the CTM Detector and improving its performance on this training corpus. Every rule modification was evaluated with respect to precision and recall, and very often changes had to be undone. It is expected however that the CTM Detector will perform similarly with other test data, as the possibilities of expressing conceptual time metaphors are limited but its applicability to other texts remains to be tested in future research. The CTM Detector could be improved by running it on other data, evaluating the new results and modifying the rules and lexica of the program. The Stanford Parser performed well in analyzing sentences. It has to be pointed out however that 51.92% of the false positives and 52.63% of the false negatives were caused by erroneous outputs of the parser. Improvements in performance could be achieved by embedding another dependency parser into the CTM Detector. In addition, Chapter 6 revealed that the CTM Detector requires components for solving lexical ambiguity (*Word-Sense Disambiguation*) and for identifying *multiword expressions* and *named entities*.

The automatic detection of conceptual metaphors is an entirely new research field with a lot of potential to improve various areas of natural language processing. In machine translation for example, the detection of conceptual metaphors is critical when an input that contains a conceptual metaphor is automatically translated to a target language in which it is non-existent. A discussion in *Universality of Conceptual Time Metaphors* (Section 2.4) presented observations that indicate that the Moving Ego metaphor is avoided in Korean. One of the examples was the English sentence (7.1) that was transformed by a professional Korean translator through Time-Model-Switching, as repeated here in (7.2).

(7.1) "Since we are approaching the Tenth Anniversary of our establishment ..."

(7.2) ENG: Since we are **approaching** the Tenth Anniversary of our establishment, ...

KOR:	창립	10주년	기념일이	다가오므로 ...
	changlib	10 junyeon	ginyeomil-i	daga-omeulo ...
	establishment	10 years	anniversary-SUBJ	close-come.as ...

"As the Tenth Anniversary of our establishment is **approaching**, ... "

One of the weaknesses of *Google Translate*[1] and similar statistical machine translation systems is that they fail to detect Moving Ego metaphors and therefore produce outputs that are unacceptable in Korean, and very probably in many other languages in the world. Example (7.3), produced by Google Translate, shows that sentence (7.1) is not returned correctly in Korean as it is translated literally, word-for-word.

[1] `http://translate.google.com/` – (Date: 29.04.2013)

(7.3) ENG: Since we are **approaching** the Tenth Anniversary of our establishment …

KOR:	우리는	우리의	설립	10	주년을	접근하고	있기	때문에
	uli-neun	uli-ui	seollib	10	junyeon-eul	jeobgeun-ha-go	iss-gi	ttaemune
	we-SUBJ-FOC	we.POSS	establishment	10	anniversary-ACC	approach-do-PROG	COP.because	because

"Because we are **approaching** the tenth anniversary of our establishment… "

(*Produced by Google Translate – Date: 29.04.2013*)

Moving Ego constructions like these could be translated correctly through a preprocessing step that involves the detection of Moving Ego metaphors and sentence modifications according to the rules of Time-Model-Switching. The first step is already accomplished by the CTM Detector which returns the sentence as (7.4).

(7.4) Since we are [approaching] the Tenth [Anniversary: TIME] of our establishment …- MOVING EGO

The development of the second module that converts the original sentence to the Moving Time metaphor could be targeted in the near future. The dependency parse in (7.5) of the CTM Detector could be used as input as it provides sufficient information about grammatical relations such as subject and object.

(7.5) **Since we are approaching the Tenth Anniversary of our establishment**, …

mark(motion_mET:approaching-4, Since-1)
nsubj(motion_mET:approaching-4, we-2)
aux(motion_mET:approaching-4, are-3)
root(ROOT-0, motion_mET:approaching-4)
det(time-7, the-5)
nn(time-7, Tenth-6)
dobj(motion_mET:approaching-4, time-7)
appos(time-7, Anniversary-9)
poss(establishment-13, our-12)
prep_of(time-7, establishment-13)
…

The Moving Time variant of the sentence could be created through replacement of the subject with the object, and the readjustment of the subject-verb agreement. Sending the Moving Time version of the original sentence through Google Translate would produce a time metaphor acceptable in all target languages, regardless of whether the Moving Ego metaphor exists or not.

The concept of time is only one of thousands of abstract concepts that are conceptualized through cross-domain mappings. Conceptual metaphors are omnipresent in everyday language as they manifest the conceptual structure of our cognitive system. Some conceptual metaphors are said to be universal and some are culture-specific. One consequence is that many sentences entered in machine translation systems will very probably contain conceptual metaphors and that these systems will cause erroneous outputs if the metaphor does not exist in the target language. The "KNOWING IS SEEING" metaphor is one of the classic examples used in the Conceptual Metaphor Theory, and it is held to be universal: "The internal self is pervasively understood in terms of the bodily external self, and is hence described by means of vocabulary drawn (either synchronically or diachronically) from the physical domain. Some aspects of the instantiation of this metaphor may be fairly common crossculturally, if not universal – for example, the connection between vision and knowledge – while others (in particular less general aspects such as the choice of the vital organ which is thought to be the seat of emotion) may vary a good deal between cultures" (Sweetser, 1990). Sweetser (1990) drew this conclusion after discovering this metaphor in many European languages. What the research is missing is a closer look at Asian languages. In Korean for example the "KNOWING IS SEEING" metaphor is avoided. Preliminary research on the English-Korean translation corpora showed that Koreans only "know" and "understand" what somebody is saying, as illustrated in examples (7.6 – 7.7).

(7.6) ENG: I see what you mean.
KOR: 무슨 말인지 알겠다
museun mal-inji al-gess-da
which word-COP understand-probably-PLA
"I probably understand the word"

(*Source: Neungyule Education*)

(7.7) ENG: I see what you guys are thinking.
KOR: 여러분이 무슨 생각을 하고 있는지 알겠어요.
yeoleobun-i museun saenggag-eul ha-go iss-neunji al-gess-eoyo
you_all-SUBJ which thought-ACC do-PROG COP-PROG know-probably-POL
"I probably know what you all are thinking."

(*Source: Neungyule Education*)

Tests on Google Translate showed that errors were also made with expression of the "KNOWING IS SEEING" metaphor. The verb "see" is always translated literally as "참조하다" which means "to see". The outputs were all unacceptable and incorrect. These conceptual metaphors have to be detected as well and transformed through "*Conceptual Metaphor Resolution*". This term refers to the replacement the word of the source domain (e.g. "see") with the equivalent word in the target domain (e.g. "know","understand"). Resolving the conceptual metaphors would produce sentences such as (7.8 – 7.9), and this transformation would raise the probability that the output of machine translation system will be correct.

(7.8) I know/understand what you mean.

(7.9) I know/understand what you guys are thinking.

These examples illustrate the great importance of processing linguistic realizations of conceptual metaphors. It is very probable that many conceptual metaphors may not be universal as predicted. For this reason, further research should be done on automatically detecting as many conceptual metaphors as possible and modifying them either through Conceptual Metaphor Resolution or through more complex transformations such as Time-Model-Switching. This thesis might be one of the first attempts to do research on conceptual metaphors within the domain of natural language processing and it has only focused on a small fraction of a larger picture. Hopefully, more and more researchers and scholars will become aware of the potential of the Conceptual Metaphor Theory to improve natural language applications in the months and years to come.

Bibliography

Aristotle, S. Halliwell, W. Fyfe, D. Russell, and D. Innes (1995). *Aristotle: in twenty-three volumes. Poetics, Longinus - On the sublime, Demetrius - On style*. Loeb classical library. Harvard University Press. 3

Cameron, L. and A. Deignan (2003). Combining large and small corpora to investigate tuning devices around metaphor in spoken discourse. *Metaphor and Symbol 18*(3), 149–160. 1, 24

Carroll, J., G. Minnen, and T. Briscoe (1999). Corpus annotation for parser evaluation. *CoRR cs.CL/9907013*. 27

Charteris-Black, J. (2004). *Corpus approaches to critical metaphor analysis*. Basingstoke: Palgrave Macmillan. 1, 24

Cho, S.-E. (2009). Models of time in Korean and English. BA Thesis (Ruhr University Bochum). 10

Chomsky, N. (1956). *Three Models for the Description of Language*. 24

Clark, H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive Development and the Acquisition of Language*. New York: Academic Press. 4

de Marneffe, M.-C., B. MacCartney, and C. D. Manning (2006). Generating typed dependency parses from phrase structure trees. In *LREC*. 25, 26, 27

Deignan, A. (2005). *Metaphor and Corpus Linguistics*. Converging evidence in language and communication research. J. Benjamins Pub. 1, 59

Evans, V. (2004). *Structure of Time: Language, Meaning, and Temporal Cognition*. Philadelphia, PA, USA: John Benjamins Publishing Company. 5, 6, 7, 8

Guerini, M., C. Strapparava, and O. Stock (2008). CORPS: A corpus of tagged political speeches for persuasive communication processing. *Journal of Information Technology Politics 5*(1), 19–32. 14

Haspelmath, M. (1997). *From space to time: temporal adverbials in the world's languages*. LINCOM Europa. 1, 4

Hudson, R. A. (1987, 03). Zwicky on heads. *Journal of Linguistics 23*(1), 109–132. 25

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, Mass.: MIT Press. 1, 4

Jewell, E. J. and F. Abate (2001). *New Oxford American Dictionary*. Oxford reference online premium. Oxford University Press, USA. 46

Kaplan, R. M. and J. Bresnan (1995). Lexical-functional grammar: A formal system for grammatical representation. 24

Klein, D. and C. D. Manning (2003). Accurate unlexicalized parsing. In E. W. Hinrichs and D. Roth (Eds.), *ACL*, pp. 423–430. ACL. 25

Lakoff, G. (1992). Metaphor and war: The metaphor system used to justify war in the Gulf. In H. Kreisler (Ed.), *Confrontation in the Gulf: University of California professors talk about the war*. Institute of International Studies, University of California. 14

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and Thought*. Cambridge University Press. 3

Lakoff, G. (1996). *Moral Politics: What Conservatives Know That Liberals Don't*. Political science/Linguistics/Cognitive science. University of Chicago Press. 14

Lakoff, G. (2005). *Moral politics. How liberals and conservatives think* (2. ed.). Chicago [u.a.]: University of Chicago Press. 14

Lakoff, G. and M. Johnson (1980). *Metaphors We Live By*. University of Chicago Press. 1, 3

Lakoff, G. and M. Johnson (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books. 4, 6, 7, 8, 9, 11

Langacker, R. (1987). *Foundations of Cognitive Grammar: Theoretical prerequisites*. Stanford University Press. 1, 4

Lin, D. (1995). A dependency-based method for evaluating broad-coverage parsers. In *IJCAI*, pp. 1420–1427. Morgan Kaufmann. 27

Màrquez, L., X. Carreras, K. C. Litkowski, and S. Stevenson (2008, June). Semantic role labeling: an introduction to the special issue. *Comput. Linguist. 34*(2), 145–159. 17

Phan, X.-H. (2006). CRFTagger: CRF English POS tagger. 24

Pollard, C. and I. Sag (1994). *Head-Driven Phrase Structure Grammar*. Studies in Contemporary Linguistics. University of Chicago Press. 24

Searle, J. R. (1993). Metaphor. In A. Ortony (Ed.), *Metaphor and Thought*. Cambridge University Press. 3

Sweetser, E. E. (1990). *From etymology to pragmatics : metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press. 60

Tesnière, L. (1959). *Éléments de syntaxe structurale*. Éditions Klinksieck. 25

Zwicky, A. M. (1985). Heads. *Journal of Linguistics 21*, 1–29. 25