

Pragmatic information extraction from subject ellipsis in informal English

Shigeko Nariyama

Asia Institute

University of Melbourne

Melbourne Australia

shigeko@unimelb.edu.au

Abstract

Subject ellipsis is one of the characteristics of informal English. The investigation of subject ellipsis in corpora thus reveals an abundance of pragmatic and extra-linguistic information associated with subject ellipsis that enhances natural language understanding. In essence, the presence of subject ellipsis conveys an ‘informal’ conversation involving 1) an informal ‘Topic’ as well as familiar/close ‘Participants’, 2) specific ‘Connotations’ that are different from the corresponding full sentences: interruptive (ending discourse coherence), polite, intimate, friendly, and less determinate implicatures. This paper also construes linguistic environments that trigger the use of subject ellipsis and resolve subject ellipsis.

1 Introduction

The interpretation of pragmatic information, such as intention, implicature¹, underspecified reference, as well as extra-linguistic information, is a prohibitively difficult task in natural language understanding (NLU) at present. This paper demonstrates that this kind of information can be extracted from a small linguistic phenomenon; that is, subject ellipsis observed in informal English.

¹ An implicature can be defined as anything that is inferred from an utterance but that is not a condition for the truth of the utterance (Levinson 1983:127).

Ellipsis has a part in the study of anaphora, as it is often referred to as ‘zero anaphora’. Anaphora resolution in English (pronominal resolution in particular, and to lesser extent bridging anaphora) has been a challenging topic for some decades both in NLU and linguistics. In contrast, the study of ellipsis in English centers around VP ellipsis and little discussion has been made on subject of nominal ellipsis (naturally because of its infrequent occurrence in formal written texts), and much less still on pragmatic effects generated from subject ellipsis. The study in NLU has approached the topics (VP ellipsis and pronominal anaphors) with the interest of resolving its referent and coreferencing, while linguistics has more concerned with the organization of discourse coherence.

The goal of this paper is to amalgamate both approaches and interests, and more importantly, to draw implicatures that are generated by subject ellipsis, by delineating various types of pragmatic information associated with subject ellipsis. First, it refutes some of the commonly held misconceptions regarding subject ellipsis (Section 2). Section 3 examines the linguistic environments (types of texts, elided subjects and predicates) that trigger subject ellipsis. It also accounts for resolving the referent of subject ellipsis, as this can cause a problem in English for not having extensive subject-verb agreement. Section 4 construes the type of implicatures that subject ellipsis gives rise to. Section 5 suggests a preliminary procedure.

2 Misconceptions about subject ellipsis

Ellipsis, and anaphora more generally, are said to play a major part in the organisation of conversation and narrative for reasons of economy, dis-

course coherence, and style (e.g. Halliday and Hasan, 1976). However, subject ellipsis seems to operate rather differently from anaphor and general ellipsis. The findings of this paper dispute or complement ‘economy’ and ‘discourse coherence’ as reasons for its use, in particular, the following commonly believed characteristics on subject ellipsis. They are that subject ellipsis is used for:

- 1) Economy, for speaking fast
- 2) Coherence
- 3) Conversation (i.e. spoken dialogue)
- 4) Spontaneous (unplanned) speech
- 5) First person pronoun elided in a declarative and second person in an interrogative

1) Regarding economy, subject ellipsis happens for more than simple reasons of economy. In fact, it is suspected that the speakers of English consciously resort to subject ellipsis when they wish to economize their utterance. If the view of economy is maintained, all subject ellipses would be spoken fast and be observed in every hasty utterance, which is not the case in the corpora. Some subjectless sentences, such as ‘(It’s) been a long time’ is often spoken slowly with underling emotions. Rather, subject ellipsis seems to be employed more for conveying different implicatures (see §4) that underlie them, that are different from the corresponding full sentences with overt subject.

2) Regarding discourse coherence, one of the implicatures given rise to by the use of subject ellipsis is ‘interruptive’ and hence it connotes the speaker’s intention to end the current topic or conversation itself (see §4.1), by which in effect discourse coherence is discontinued.

3) The perception that subject ellipsis is found in conversation is inaccurate, as it is also found in many informal written texts, such as diaries, postcards, emails, logs and blogs on the internet.

4) Subject ellipsis is believed to be a result of informal spontaneous utterances without planning and editing, but in fact it is not limited to this type. For example, the automated teller machine of a bank in New Zealand gives out this message at the end of inputting instructions, ‘OK. Got that.’ This is by no means an unplanned message; in fact it is a prudently planned message for a particular implicature, i.e. to make the response sound friendly.

5) The last misconception, which relates to ellipsis resolution, is that subject ellipsis in English

operates on the principle that first person pronoun is elided in a declarative and second person in an interrogative with a rising interrogative intonation, as in "(I) got in late." versus "(You) got in late?" This certainly plays a part in the mechanism of subject ellipsis, and is probably true across languages. However, the corpus analyses found that while first person subject ellipsis is prominent, second person subject ellipsis is rare (see §3.2)

3 Linguistic conditions for subject ellipsis

A missing subject in English² is syntactically prominent and hence is relatively easy to mechanically detect. However, for speech data it poses an issue. Some subject ellipses can be marginal between ellipsis and phonetic erosion (e.g. inaudible subject word being spoken *sotto voce* and quickly), and an appropriate heuristically oriented threshold based on instrumental analysis (e.g. the use of spectrography) must be set up as a criterion for determining ellipsis, although across-the-board ‘acoustic correlates’ of the subject ellipsis may remain as an equivocal issue. For simplicity and reasons that the main theme of this paper is not the determination of ellipsis itself, this paper considers the transcriptions of spoken texts. Moreover, notwithstanding that the role of prosodic features plays a significant part in determining implicatures, given the scope of this paper, it is basically put aside for future study. Nonetheless, the findings are still meaningful for analyzing informal utterances in the internet domains, such as emails, logs and blogs, the type of text that is increasing exponentially in quantity and importance in NLU particularly for being able to process mechanically.

It is conceded that the description of subject ellipsis in this paper is drawn from small corpora in NLU standards, although covering various types of texts and studies in the literature, so that the findings from this analysis are intended to provide a starting point in the study of the pragmatic information that subject ellipsis prevails.

² Subject ellipsis refers to those unexpressed subjects occurring in the sentence initial position, hence excluding subject ellipsis in coordinate structures and non-finite clauses (adverbial clauses, gerundive clauses, and prepositional clauses). This is drawn from the claim that non-syntactically motivated subject ellipsis only occurs in the sentence initial position (McShane 2005, Nariyama 2004, Swales 2002, Cote 1996), except that Haegeman and Ihsane (1999) claim that a diary may exhibit some use of subject ellipsis in embedded clauses.

3.1 Type of texts

Undoubtedly subject ellipsis is a feature of informal register. We would not expect to hear subjectless sentences, for example, at court hearings. Indeed, Swales (2002) examined a corpus of academic speech (MICASE), consisting of 36 speech events, covering such texts as colloquia, dissertation defenses and lab meetings. He reports that subject ellipsis is rare in formal speech; the highest ratio was found for *DEPENDS* at 14%, followed by *LOOKS* at 8.1%.

Swales also notes the dialectal aspect that subject ellipsis is a more prominent feature for British English (also by Cote 1996); the rate of subject ellipsis before *DEPENDS* is 60% in British English conversation and 30% in American English. Comparing these with 14% in MICASE, subject ellipsis is indeed a character of informal speech.

Further evidence is found in Taylor (2002) in analysing Australian English Corpus (Monash University 1996~1998) that *Gotta* often occurred without an overt subject, all with first person subjects. The contracted forms of lexicon, such *gonna*, *gotta* and *hafta* (so called ‘the quasi-modals’), are undoubtedly representative of casual speech.³

However, what may not be so obvious is that an informal register does not license the use of subject ellipsis at all times. The corpus analysis on TV drama scripts (§3.2) found that subject ellipsis had a propensity to emerge at particular scenes and topics where the atmosphere of the scenes is both casual and friendly, while it did not occur with the same speech participants at other scenes and topics. This is plausible from the fact that when asking a favour or showing gratitude, more formal language (i.e. without ellipsis) tends to be used even to close participants. Or colleagues may frequently say “Dunno” and “Doesn’t matter” to each other. However, in a situation where one lost his job, he is unlikely to use subjectless sentences in response to a question “What happened?”

3.2 Type of elided referents

Table 1 shows the results of the referent distribution of subject ellipsis from analyzing three cor-

³ The use of subject ellipsis in non-informal texts is also found (Cote 1996). Recipe and instruction texts constantly elide subject (as well as object). This usage is domain specific and the referent is fixed as a second person subject.

pora: 1) three transcripts of family conversation (FaCon) drawn from Australian English Corpus (Monash University 1996~1998) collecting family interviews about their past holidays (Nariyama 2004); 2) three 30-minute-TV Australian drama transcripts (TV) (Nariyama 2004); 3) Switchboard corpus consisting of telephone conversation on a variety of specified every day topics (Cote 1996).

Referent	FaCon	TV dramas	Switchboard
I	10 (20.4%)	25 (47.2%)	47 (26.0%)
we	2 (4.1%)	0	3 (1.7%)
you	4 (8.2%)	3 (5.6%)	13 (7.2%)
he/she	2 (4.1%)	6 (11.3%)	47 (25.9%)
it	30 (61.2%)	17 (32.1%)	67 (37.0%)
they	1 (2.0%)	2 (3.8%)	4 (2.2%)
Total	49 (100%)	53 (100%)	181 (100%)

Table 1: Type of elided referents and their frequency of occurrence by type of texts

The distributions of ‘I’ and ‘it’ are different among the texts; more ‘it’ and less ‘I’ in the two corpora than in TV. This is attributed to the fact that the data were drawn from conversations with particular topics rather than a free conversation. Hence many utterances relate to the past mentions of descriptions where the anaphoric use of ‘it/he/she’ is more relevant (see Table 2 §3.4.1 for the ratio). Nonetheless, what is common among the three is that 1) first person pronoun ellipsis is frequent, 2) second person ellipsis is rare, although the corpora are a collection of dialogues (§3.3).

3.3 Informativeness

Semantic informativeness of a sentence plays a major part in enabling subject ellipsis. It is manifested in two ways: type of verbs & adjectives and the amount of information expressed in a sentence.

3.3.1 Type of verbs and adjectives

Subject ellipsis has a strong association with, and hence is triggered by, particular verbs and adjectives (with or without auxiliary). Swales (2002) found the following verbs to be of such types in MICASE (academic speech corpus): *DEPENDS*, *SEEMS*, *SOUNDS*, *LOOKS*, *TURNS OUT*, *WANNA*, *HEARD*, *SEEN*, and *GOT*.

This does not, however, explain the frequency of first person subject ellipsis and rareness of second person subject ellipsis seen in Table 1. Nariyama

(2004) claims that it is fundamentally an epistemic reasoning (having sufficient knowledge about a statement) that has substantial controls on the application of subject ellipsis. The most of the verbs above that come with subject ellipsis require epistemic knowledge and hence first person subject ellipsis. Even when the subject is ‘it’, the agent (psychological subject) is still first person.

Because of epistemic reasoning, semantically rich and private verbal lexicon can only be used for first person subject, as in the (a) examples, and this makes second person subjectless sentences unacceptable even in an interrogative, as in the (b)s.

- (1a) (I’d) love a coffee.
 (1b) * (Would you) love a coffee?
 (2a) (I’m) feeling fantastic.
 (2b) (*) (Are you) feeling fantastic?
 (3a) (I) wouldn’t mind a coffee.
 (3b) (*) (You) wouldn’t mind a coffee?

‘Love’ conveys high degree of preference as well as a request, which is privy and subjective to the speaker. Thus, the speaker can state his own feeling, as well as make a request, but cannot do so for others;⁴ hence (1b) is unacceptable; analogously for (2b) and (3b), or ‘hate’, ‘thought’, and ‘hope’. This unacceptability of sentences with second person subject often remains even with overt subjects.

3.3.2 Amount of information

Analogous to the richness and privy of lexicon, informativeness in terms of amount of information (number of phrases) governs the acceptability of subject ellipsis. It is less restricted for sentences with first person subject as in the (a) examples than non-first person subject as in the (b)s.

- (4a) (I) had a good time.
 (4b) (Did you) have a good time?
 (5a) (I) had a wonderful time.
 (5b) (?) (Did you) have a wonderful time?
 (6a) (I) had a good time visiting my family in Sydney last week.
 (6b) (*) (Did you) have a good time visiting your family in Sydney last week?

⁴ Japanese is well documented for having rigid constraints on subjective statements (Nariyama 2003, Aoki 1986, inter alia).

3.4 Referent resolution

The recoverability of referent is the imperative condition on the employment of ellipsis, but English has limited subject-verb agreement. So, how is the referent of elided subject retrieved?

3.4.1 Locational constraint

The results in Table 1 are further examined by the type of subject ellipsis, and are summarized as follows (mostly drawn on the TV corpus):

- 1) Anaphoric (23/53)
- 2) Deictic (21/53)
- 3) Idiomatic (conventionalised usage) (8/53)
- 4) Expletive ‘it’ (1/17, c.f. FaCon:2/30)

The corpus analysis concludes that 2) Deictic, 3) Idiomatic (§3.4.2) and 4) Expletive subject ellipses can occur freely, but 1) Anaphoric has a locational restriction to occur immediately after the sentence with the referent, for example:

- A: Where’s dad?
 B: [TV33] (He’s) birthday shopping, I bet.
 [TV34] (He’s) so bad at pretending.

Where the referent is not found in the immediately preceding sentence, quasi-right dislocation was used to express the subject at the end of the sentence, presumably to ensure that no ambiguity of reference would occur, for instance:

Quasi-right dislocation

- A: He’s going to do a Thorpey after today.
 B: Why? What happened?
 A: [TV24] Won the year 7 freestyle, he did.

Although the use of ‘it’ was frequent in Table 1, the expletive use was rare (See Lappin and Leass (1994) for identifying the expletive use of ‘it’.) As mentioned, this is attributed to the type of corpora having specific topics of conversation.

Function of ellipsis	Number (FaCon)	Number (TV)
expletive	2	1
deictic	0	1
anaphoric	28	15
Total	30/49 (61.2%)	17/53 (32.1%)

Table 2: Frequency of the function of ‘it’

3.4.2 Complementary distribution

The constraints on informativeness described in §3.3 often produce complementary distribution with regard to the type referent for subject ellipsis, and hence they in turn signal the referent of ellipsis. For example, ‘love’ is associated with first person subject (7a) and ‘like’ with second (8a):

(7a) (I’d) **love a coffee.**

(7b) *(I’d) like a coffee.

(8a) (Would you) **like a coffee?**

(8b) *(Would you) love a coffee?

Likewise, idiomatic subjectless expressions are essentially set phrases whose meanings are self-contained in their own right, so that elements in the expressions tend to be fixed in terms of person, the declarative/interrogative, polarity, and verbs. Hence, these constraints resolves the referent of subject ellipsis. For example, ‘gonna be long’ is strongly associated with second person subject and therefore occurs in interrogatives as in (9a). It sounds odd to be used for first person subject even in a declarative (9b) or with the full sentence (9c). Instead (10a) is likely to be used for the proposition, which in turn is awkward for a second person subject (10b). Thus, it creates complementary distribution.

(9a) (Are you) **gonna be long?**

(9b) *(I’m) gonna to be long.

(9c) ? I’m gonna to be long.

(10a) (I’m) **gonna be a while.**

(10b) (*) (Are you) gonna be a while?

Analogously, ‘won’t be a minute’ is strongly associated with first person in declarative (11a), so that the same with second person (11b) is unacceptable unless it is quoting or parodying the earlier statement. The same goes for ‘just be a minute’ in (12a) and (12b).

(11a) (I) won’t be a minute.

(11b) * (You) won’t be a minute?

(12a) (I’ll) just be a minute.

(12b) * (You will) just be a minute?

Polarity is also often fixed for a particular expression. For example, (13a) is set for first person

subject and negative, so that any variation to it gives rise to an unacceptable sentence; e.g. second person negative interrogative (13b) (unless quoting or parodying what the person has just said), first person non-negative declarative (13c), and second person non-negative interrogative (13d) are all unacceptable. Interestingly, (13d) will be acceptable if it is expressed slightly differently with an overt subject, as in “Do you mind having a coffee (because I haven’t got anything else at the moment)?”

(13a) (I) wouldn’t mind a coffee.

(13b) * (You) wouldn’t mind a coffee?

(13c) * (I) mind a coffee.

(13d) * (You) mind a coffee?

Thus, the constraints on person, declarative/interrogative form, polarity, and verbal semantics in turn allow little ambiguity in recovering the referent of subject ellipsis.

3.4.3 Discerning subject ellipsis from imperatives

Subject ellipsis in the initial position is structurally identical with the imperative construction in English. However, potential ambiguities in the interpretation are resolved by a number of factors. First, the semantic content of privy verbs cannot be forced upon someone else, and hence does not make sense for such subjectless sentences to be interpreted as imperatives, for instance:

(14) * Like a coffee!

c.f. (14a) Have a coffee!

(15) * Feel alright!

The second is the tense complementarity. Some verbs tend to occur in a past tense, e.g. ‘got’, while imperatives do not allow past tense:

(16a) * Got a house!

(16b) Get a house!

The third is that the presence and type of the object can make the distinction; e.g. “Tell you” will be interpreted as a subjectless sentence “(I’ll) tell you”, while “Tell me” will be interpreted as an imperative “(You) tell me”. Analogously, “See you later” is interpreted as a subjectless sentence “(I’ll) see you later”, while “See me later” as an imperative/request sentence “(You) see me later.”

4 Five basic implicatures

Undoubtedly intonation and context play a large part in the inferred meanings. Nonetheless, the iconicity in the implicatures is observed in the following sets of examples. (a)s with subject ellipsis convey specific meanings without being linguistically expressed in a strict sense, while (b)s convey unmarked linguistic meaning.

- (17a) (I've) gotta go.
(17b) I've got to go.
- (18a) (I) dunno.
(18b) I don't know.
- (19a) (I've) got it.
(19b) I've got it.
- (20a) (It) doesn't matter.
(20b) It doesn't matter.
- (21a) (I) should've known better.
(21b) I should have known better.
- (22a) (It's) been a long time.
(22b) It's been a long time.

(17a): a more evasive and dismissive motive,
(17b): tends to imply a more honest/genuine statement that the speaker actually has to get to a particular place by certain time.
(18a): an indeterminate state of mind such as 'I'm not sure', 'I haven't thought about it', or a dismissive motive, such as 'I don't want to think about it.', or even to the extent, 'I don't care'.
(18b): more genuine: 'I thought about it, but I have no idea.'
(19a): more emphatic and the speaker may have anticipated what the interlocutor has just said and therefore (19a) is a little hasty-sounding.
(20a): more likely about trivial matters
(20b): for more important, serious matters, and for giving consolation.
(21a): less directed/emphatic and therefore less punitive and apologetic that the mistake is understandable, silly, or trivial.
(22a): has a restricted usage directed at someone intimate to the speaker.

Five basic implicatures can be drawn from the above and the earlier examples: 1) Interruptive with dismissive/evasive motives (e.g. all of the

above except 22a), 2) Polite (e.g. 21a), 3) Intimate (e.g. 22a), 4) Friendly (e.g. 1a, 9a, 19a), and 5) Less determinate implicatures (e.g. 18a, 'Depends', 'Seems').

Subject seems to drop for two basic reasons. First, for 1), 4) and 5) implicatures, the meaning conveyed by subjectless utterances tends to be evasive, less determinate (informative, definite, formal), and spoken fast. It is a logical tendency for semantically insignificant elements to be unstressed, and unstressed pronouns drop, which is economical in conversation. The other is that the subject is intentionally underspecified either for disguising the identity for 2), or for the effect that the absence of the subject makes the identity conspicuous for 3).

4.1 Interruptive implicatures with dismissive/evasive motives

Speech act participants generally have the intention to converse with one another and sustain their conversation, as described in Grice's hallmark discovery of the Cooperative principle (1975). While full sentences to some degree elicit responses from the addressees and therefore aid conversation flow, the corresponding subjectless sentences tend to convey to the addressee implicatures of fulfilling social obligation, keeping a low conversational profile, and minimising invitation of response to the subjectless utterance. For example, "(I've) gotta go" tend to imply that the speaker is fulfilling his social obligation by acknowledging the presence of the addressee in making an utterance, but at the same time indicating that he is not inviting any meaningful response. Indeed, this example was used in the TV drama instead of saying "Good-bye".

If this view is maintained, subject ellipsis can have the effect of changing a topic or ending a conversation itself, in which case it has the opposite effect to what has been claimed on anaphora and ellipsis; anaphora is one means of establishing coherence (e.g. Fox 1996, Halliday and Hasan 1976). This claim is plausible that the use of ellipsis does create cognitive states of coherence on the grounds that the addressee has to look elsewhere for the interpretation of the missing subject; and in doing so links the current sentence to another sentence. It is certainly true for anaphoric use of subjectless sentences, but for the above usages it is questionable.

4.2 Polite implicatures

Some subjects seem deliberately unspecified in order to conceal the subject identity or make it ambiguous for politeness; for example. (Note that in (23) concord is also taken out.)

- (23) (I/you/we/he/they/.. 've/s) got to have a coffee.
(24) (I/you/we/he/they/...) should've known better.
(25) (I/you/he/they/...) haven't/hasn't got a chance.
(26) (I/you/we/he/they/...) should send a postcard.

The subject of (23) can be interpreted as anyone in the presence of a group of people, but the intended referent may in fact be 'I' because 'I'm tired', 'you' because 'you look sleepy', or 'we' because 'we all worked so hard'. This under-specified subject avoids a direct speech act and one's responsibility/accusation/self-centeredness (e.g. 'I got to have a coffee'), and softens the implicatures by creating an indirect request/suggestion. Even when the intended referent is clear, it is left up to the addressees whether or not to interpret the elided subject as being directed to himself. This indirectness is seen as one type of politeness strategy (Brown and Levinson 1987, Leech 1983).

(27b) is an interesting example from the TV corpus for having the first person subject ellipsis instead of second in an interrogative '(Would you) like another?', which may contribute to the rareness of second person subject. The same speaker uttered (27a) and (27b) with no pause in between.

- (27a) (It) looks cold. [The speaker is looking at
the addressee's cup of tea.]
(27b) (Shall I) **make another** (cuppa)?

While the semantic content of the sentences is virtually identical, some speakers of English find (27b) more polite than the one with the second person subject, because it offers help as well as asks the desire of the addressee. This view is consistent with the description by Leech (1983) that the more benefit an utterance brings to the hearer, the more polite it is. Interestingly, (27b) cannot be interpreted as '(Would you) make another?' in the context.

4.3 Intimate implicatures

Some subjectless sentences convey intimate implicatures, when spoken slowly with specific prosody.

- (28) (It's) been a long time.
(29) (It's) nice to see you.
(30) (It's) lovely to get your email.

4.4 Friendly implicatures

Subjectless sentences with friendly implicatures relate to particular expressions, such as 'Had a good time' and 'Like a coffee?'

Friendly and Intimate implicatures overlap somewhat, but they differ with the view that the subject in the former drops for being friendly and therefore causal, while in the latter the absence of the subject makes the obvious identity of the subject conspicuous, more meaningful, and private.

4.5 Less determinate implicatures

Owing to the semantics of the verbs, subjectless sentences with such verbs as *DEPENDS*, *SEEMS*, *SOUNDS*, *LOOKS*, *TURNS OUT*, *HEARD*, and *SEEN* tend to convey less determinate, definite, less objective implicatures. Arnold Zwicky makes similar notes (Language log March 19 2005), in that 'Odd that Mary never showed up' is more expressive/subjective, while the corresponding full sentence 'It is odd that Mary never showed up' is more reportive/objective.

4.6 Discriminating the implicatures

All things in pragmatics are extremely convoluted involving seemingly limitless factors that influence the ultimate interpretation, and worse still all interpretations are defeasible. Thus, the following is a rough description of features that discriminate the five implicatures at this stage.

Each implicature relates to particular expressions deriving from their lexical semantics, particularly the polite and intimate implicatures. Some expressions such as 'Got it' can have multiple implicatures. The qualitative studies with prosodic features will reveal the type of expressions associated with each implicature, for example, the fast spoken 'Got it' is to imply interruptive implicature.

Then the derived implicatures can in turn signal, for example, that the polite implicatures are associated with multiple interpretations of the subject identity (in which case ellipsis resolution is less important for this type of ellipsis), or that the intimate implicatures reveal the speaker's loving relationship with the addressee.

5 Resolution procedure

A preliminary rough procedure regarding subject ellipsis in English is briefly suggested here.

1. *Detect* a sentence-initial subject ellipsis.
 - 1.1 Discard if it is an imperative based on §3.4.3.
2. *Pragmatic information extraction*
Draw an implicature based on §4.6 with consideration to the dialectal differences (§3.1).
3. *Extra linguistic information extraction*
Subjectless sentences imply ‘informal’ conversation involving an informal ‘topic’ and familiar/close ‘participants’.

6 Conclusion

This paper has made a wide description concerning subject ellipsis in English: the linguistic environments that trigger subject ellipsis, including the subject ellipsis resolution, and the marked implicatures conveyed by subject ellipsis that are different from those given by the corresponding full sentences with overt subject. It seems paradoxical that less words, i.e. more ellipses, convey more internal feelings and intentions.

Subject ellipsis, and elliptical constructions more generally, are an essential feature of everyday conversation and are a common phenomenon cross-linguistically. Gilligan (1987) reported, based on a sample of 100 languages, that only seven of these do not allow subject ellipsis in finite clauses. Since English is not generally known for ellipsis, the study on other languages may reveal more interesting outcomes.

Finally, the work on this topic is in its infancy and great more work ahead of us before the findings can be put to meaningful use in an NLU system. First, a quantitative investigation with prosodic features is a must for assuring the findings of this paper and creating an inventory of subject ellipsis. Further investigation will point to the direction for the most appropriate applications and methods to implement this sort of pragmatic information into a system. The use of findings may be more accurate and feasible for generation than for understanding. Nonetheless, it will be of use to applications that are in search of contextual cues to identifying the type of topic, participants and their relationship, and in rephrasing relatively formal sentences to more informal ones.

References

- Haruo Aoki. 1986. Evidentials in Japanese. In Chafe & Nichols (eds). *Evidentiality: the linguistic coding of epistemology*. Norwood: Ablex 223-238
- Penelope Brown and Stephen Levinson. 1987. *Politeness: some universals in language usage*. Cambridge: Cambridge University Press.
- Cote Sharon. 1996. *Grammatical and Discourse Properties of Null Arguments in English*, Ph.D. Dissertation, University of Pennsylvania.
- Barbara Fox. 1996. *Studies in anaphora*. Typological studies in language 33. Amsterdam: John Benjamins.
- Gary Gilligan. 1987. *A cross-linguistic approach to the pro-drop parameter*, University of Southern California Dissertation.
- Paul H. Grice. 1975. Logic and conversation. In P. Cole & J.L. Morgan eds., *Syntax and semantics 3: Speech Acts*. 43-58, New York: Academic Press.
- M.S.K. Halliday and R. Hasan, 1976. *Cohesion in English*. London: Longman.
- Shalom Lappin and Herbert Leass. 1994. An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20(4): 535-561.
- Geoffrey Leech. 1983. *Principle of pragmatics*. London: Longman.
- Stephen Levinson. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- Liliane Haegeman and Tabea Ihsane. 1999. Subject ellipsis in embedded clauses in English, *English Language and Linguistics*, 3(1): 117-145.
- Marjorie McShane. 2005. *A theory of ellipsis*. Oxford University Press.
- Shigeko Nariyama. 2003. *Ellipsis and Reference-tracking in Japanese*. Studies in Language Companion Series 66, Amsterdam: John Benjamins
- Shigeko Nariyama. 2004. Subject ellipsis in English, *Journal of Pragmatics*, 36(2): 237-264.
- John Swales. 2002. “Any last minute thoughts on this particular search?” The occurrence of sentence-initial ellipsis in research speech. In Proc. of the 4th North American Symposium on corpus linguistics and language teaching.
- Jo Taylor. 2002. The quasi-modals *gonna*, *gotta* and *hafta* in Australian English. paper presented at Australian Linguistic Society Annual Meeting at Macquarie University Sydney.