INFORMATION PACKAGING
IN JAPANESE NEWS LEAD SENTENCES

*Andreea SION[1]
andreeasion@yahoo.com

Abstract: News stories are reports about recent events or happenings that must efficiently convey accurate, clear and unambiguous information to a mass audience, in a short amount of time (if broadcasted) or space (if written). This study aims to analyze, from a discursive standpoint, the way in which information is conveyed in the news first paragraph, usually named "lead", in terms of information packaging (as defined by Prince 1981) and referent identifiability (as defined by Lambrecht 1994), simultaneously taking into account several factors, such as the type of media (broadcast /vs./ written news), the type of grounding provided for the subject noun phrases found in the lead sentences and the way the subject noun phrases are linguistically marked (as topics or non-topics).

Keywords: information packaging, identifiability, grounding, discourse topic, wa, ga

INTRODUCTION

News reports are a means of mass communication, conveying information about recent events or happenings of general interest. They must be generally accessible to a simultaneous mass audience that is also fragmented (as the individuals forming the mass audience have various educational backgrounds, interests, beliefs etc) and that does not provide immediate feedback (as would happen, for instance, in conversation) (Bell 1991:2). Simultaneously, news reports are subject to length limitations. For all these reasons, in order to successfully and efficiently deliver their information content, news reports must be concise and easy to understand, while providing all important information about the events they are reporting.

The most common method for writing news stories is the so-called *inverted pyramid*, in which the most important information is conveyed first, followed by the secondary information, given in a decreasing order of importance. The key element of news stories is the first (= *lead*) paragraph, where the information conveyed answers the 6 basic questions of news reporting, generally known as 5W1H: *who*, *when*, *where*, *what*, *why*, and *how*. The lead paragraph plays a very important role in the news discourse comprehension as it represents the starting point of communication and a base for the efficient processing of the subsequent information.

---

[1] ***Andreea SION,** Lecturer PhD (doctorate degree received from Tsukuba University, Japan)**,** Hyperion University, Bucharest.

DISCOURSE PROCESSING: BASIC THEORETICAL ASSUMPTIONS

In the past few decades, the mechanisms of discourse processing have been analyzed from a variety of perspectives, but most models seem to agree upon the following basic assumptions:

Comprehension is a dynamic process, in which information is transmitted and received in a linear way. Thus, the receiver assimilates the information incrementally, integrating it in his cognitive model (mental representation) as it is transmitted (Van Dijk & Kintsch 1983, Langston & Trabasso 1999, among others). However, because the receiver's working memory has a limited capacity, it is able to keep active only relevant information, which is the information that is strongly connected in the receiver's cognitive model, while less important information is deactivated and stored in the receiver's peripheral memory, to be reactivated when needed. For an efficient processing of new information so that it can be successfully integrated in the receiver's model, a good *anchoring* (or *grounding*) of new information in the receiver's preexistent cognitive model is essential.

The notion of *grounding* is strongly connected with that of *information packaging*, defined by Prince as follows: "the tailoring of an utterance by a sender to meet the particular needs of the intended receiver […] reflect[ing] the sender's hypotheses about the receiver's assumptions and beliefs" (Prince 1981:224).

CURRENT ANALYSIS

Having in mind the notions stated above, this study aims to analyze **a)** the types of subject noun phrase referents that may appear in Japanese news lead sentences, in terms of cognitive *activation* in the receiver's mind, **b)** the linguistic markers they take when first introduced into the discourse (i.e, as topics or non/topics) and **c)** the types of *grounding* provided for these referents.

The corpus used for the analysis consists of the lead sentences of 261 television news items and 47 newspaper news items[2] in which the subject noun phrase is marked either with *ga*, the typical nominative case particle, or with *wa*, the topic particle. Often, the lead sentences may consist of several coordinated phrases containing subject NPs marked with *ga* or *wa*, such as example (1) below, in which cases they are all taken into account.

(1)　　今日午後、横浜市の金沢区で、下校途中の小学生の列に乗用車が [subject NP1] 突っ込み、／　　1人が[subject NP 2] 死亡、／　　4人が[subject NP 3] 重軽傷を負いました。

The total number of occurrences is as shown in Figure (1) below:

---

[2] Selected out of a larger corpus of 300 television and 50 newspaper news items, after eliminating those that contain other means of expression than *ga* or *wa*.

	TV news	Newspaper news
subject NP + *wa*	149 (39.31%)	56 (50.45%)
subject NP + *ga*	230 (60.68%)	55 (49.54%)
total	379 (100%)	111 (100%)

TYPES OF SUBJECT NP REFERENTS IN THE JAPANESE NEWS LEADS

The current analysis is based on Lambrecht's *identifiability model*. According to Lambrecht, an identifiable referent is "(a referent) for which a shared representation already exists in the speaker's and hearer's mind at the time of utterance" (1994:77). In short, *identifiable* referents are referents that already are in an *active* state in the hearer's mind, or are *accessible* because they have just been activated (due to the textual or situational context), or are *inactive* (they do exist in the hearer's long/term memory, but not in an active state). *Unidentifiable* referents, on the other hand, represent brand-new information for the hearer, and may be *anchored* (linked, by means of another NP, to the hearer's mental representation) or *unanchored*.

Following Lambrecht's model, in Sion (2003) we analyzed the types of subject NP referents that appear in the Japanese news leads and their degree of "identifiability", starting from the assumption that the receivers *know* they are exposed to news reporting (in the larger context of the television news broadcast or the newspaper) and have already activated their mental "news" framework, thus creating a series of expectations about the kind of information they are about to receive.

The subject NP referents that appear in the Japanese news leads can be classified as follows:

A. Referents (assumed to be) known / familiar to the receiver at the moment they are introduced into the discourse, such as

1. Widely known national and international political/economical etc. institutions, political personalities etc. – ex. the Imperial Household, the Japanese government, foreign governments/presidents, which we consider to be [identifiable, active];

2. Japanese national political/economical/social organizations, committees, large foreign organizations etc., which we consider to be [identifiable], but only [semi-active];

B. Referents (assumed to be) unknown to the receiver at the moment they are introduced into the discourse – ex. individuals, things, events, which are

3. identifiable because they have already been introduced in the preceding context, or

4. identifiable because they are inferable from the preceding context, or

5. new, but grounded, at least to some extent, into the receiver's cognitive model, through the preceding context.

Examples:
A-1   Identifiable – active
(2)
ニューヨークを訪れている<u>**小泉総理大臣**</u>は、日本時間の明日朝、ブッシュ大統領と会談します。 "Prime-Minister Koizumi"
(3)
<u>**東京都**</u>は、国から地方に移す税に、消費税などを加えて、国税と地方税の比率を現在の3対2から1対1にすべきだなどとする、独自の案をまとめました。"Tōkyō Metropolis"

A-2   Identifiable – semi-active
(4)
<u>**みずほフィナンシャルグループ**</u>は、今日、年間で最も多い、1200万件もの口座振替や給与振替などの処理が集中しましたが、混乱なく、大半の処理を終えたとしています。
"Mizuho Financial Group"
(5)
<u>**連合**</u>は、今日の中央執行委員会で、雇用対策を盛り込んだ補正予算の編成と共に、いわゆるサービス残業をなくすことと、パートタイマーの均等待遇を確保する法律の設定を、この秋の最重点課題に位置付け、実現を目指していく方針を決めました。"The Japanese Trade Union Confederation"[3]

B-3   Identifiable – introduced in the preceding context
(6)
今年6月、埼玉県春日部市の定時制高校の生徒が死亡した集団暴行事件で、刑事処分が相当だとして、障害致死の罪で起訴された<u>16歳の少年二人</u>の初公判が開かれ、<u>**二人は**</u>起訴事実を認め、一生を掛けて償いたいと述べました。*Futari* – "the two persons" has been introduced in the preceding context, but not as a subject NP.

B-4   Identifiable – inferable from the preceding context

---

[3] The full Japanese name of the confederation is「日本労働組合総連合会」. The use of the abbreviated version,「連合」, is also a sign that the referent is considered to be identifiable, and at least semi-active, for the receivers.

(7)
東京町田市の<u>小学校</u>では、地域の人と一緒に育ててきた稲が実りまして、今日<u>子供達</u>が稲刈りをしました。The subject NP *kodomotachi* – "the children" can be inferred from the cognitive frame of *shōgakkō*, "primary school", already mentioned in the previous context.

B-5   New – with grounding context
(8)
埼玉県栗橋町で、金融機関から自転車で帰宅する途中のお年寄りから現金をひったくったとして、**31歳の男**が逮捕されました。"A 31-year old man"

The frequency of these types of referents in television and newspaper news leads is as shown in Figures (2) and (3) below:

Fig. (3) Subject NP referent identifiability in Japanese television news leads

Types of referents		Frequency		
		wa	*ga*	
A. Known referents	1. active	**58**	3	61 (16.09%)
	2. semi-active	**55**	21	76 (20.05%)
B. Unknown referents	3. identifiable – previously introduced	17	13	30 (7.91%)
	4. identifiable – inferable	10	17	27 (7.12%)
	5. new, grounded	9	**176**	185 (48.81%)
	Total	149 (39.31%)	230 (60.68%)	379 (100%)

Fig. (4) Subject NP referent identifiability in Japanese newspaper news leads

Types of referents		Frequency		
		wa	*ga*	
A. Known referents	1. active	**18**	0	18 (16.21%)
	2. semi-active	**13**	7	20 (18.01%)

B. Unknown referents	3. identifiable – previously introduced	12	2	14 (12.61%)
	4. identifiable – inferable	11	4	15 (13.51%)
	5. new, grounded	2	**42**	44 (39.63%)
Total		56 (50.45%)	55 (49.54%)	111 (100%)

In both television and newspaper news, most subject NP referents introduced in the lead sentence are, not surprisingly, referents of the B-5 type, [new/unidentifiable]. Active and semi-active referents tend to be marked with *wa* and presented as topics from their first occurrence in the discourse, while new, unidentifiable referents are marked with *ga*, not acquiring the status of topic. The rare exceptions (active referents marked with *ga* and presented as non-topics, or new referents presented as topics), as illustrated in examples (9) and (10), can be explained from a pragmatic perspective:

(9)
<u>旧第一勧業銀行</u>がおよそ１６０億円の税金を納めていませんでした。"Former Daichi Kangyō Bank"
(10)
先月首都圏で発売された**新築マンション**は、４７００個あまりで、マンション各社が大量の新規物件が出回る、秋の商戦を前に、在庫を積極的に売りに出したことから、８月としては過去最高になりました。"Newly built apartment buildings"

In example (9), "Former Daichi Kangyō Bank", though semi-active, is not marked with *wa* in order to be presented as a topic and made "what the rest of the sentence is about"[4]. The referent takes *ga* because, in this particular case, the entire lead sentence functions as an *event-reporting sentence*[5], presenting the event ("The bank did not pay its taxes") as whole, new information.

In example (10), the NP referent "Newly built apartment buildings" cannot be inferred from the previous context ("sold last month in the Tokyo Metropolitan Area) and is most certainly not activated or it is at least semi-active in the

---

[4] According to Lambrecht (1994:131), "A referent is interpreted as the topic of a proposition if in a given situation the proposition is construed as being about this referent, i.e. as expressing information which is relevant to and which increases the addressee's knowledge of this referent".

[5] A more detailed discussion on event-reporting (thetic) sentences /vs./ topic-comment (categorical) sentences is found in Lambrecht 1994:143ff.

receiver's mind. However, presenting it as such and marking it with the topic marker *wa* from its first occurrence is an economical way of conveying information by skipping the stage where the existence of the referent is stated, which also ensures a smoother discourse progression (instead of saying, for instance, 「首都圏で新しいマンションが建てられ、このうち先月発売されたものの数は…」 "New apartments have been built in the Tokyo Metropolitan Area. The number of those sold last month…").

GROUNDING

In principle, when a news item starts being broadcasted, or when the receiver starts reading a news story, he cannot really anticipate what kind of referents will be introduced into the discourse. Even in the case of identifiable referents, when they are at their first occurrence, the receiver makes a somewhat greater effort to process the information and find its proper place in his mental discourse model. In order to lessen the receiver's cognitive effort, newly introduced referents must be accompanied by some kind of additional information that would help link them to the receiver's background knowledge. For this reason, very often, before the main event is transmitted, the lead paragraph provides information on the place and time of the main event, as well as on other secondary information.

In the news corpus that we analyzed, the information provided in the lead paragraph at the time of introducing new referents, can be divided into two large types:

(i) information that *links* the new, unidentifiable event to some other event that is assumed to be already known by the receiver (in many cases, a previous event that has already been discussed by news items) - which we will call *connective grounding*

(ii) information that explains or gives details about the newly-introduced event, without linking it to some other event – which we will call *explanatory grounding*

In some cases, the referents are introduced at the very beginning of the news lead (zero grounding).

Examples:
Zero grounding
(11) **皇太子様は**今日、ヨーロッパ三ヶ国の訪問に出発されました。
(12)
小泉総理大臣は今日、北海道を訪れ、コンピュータを使った酪農経営の現場や、大学や企業などが連携して、新たな産業の育成を目指す取り組みなどを視察しました。

Connective grounding

29

(13)
<u>東京電力が、原子力発電所のトラブルを隠していた問題を受けて、新潟県</u>の**平山知事**と柏崎市の**西川市長**、それに刈羽村の**品田村長は**会談し、柏崎刈羽原子力発電所3号機のプルサーマル計画の事前了解を取り消すことを決めました。- the main event of the news, the fact that three administrative bodies had a discussion and took a decision, is linked to a previous event, an incident of not reporting malfunctions at a nuclear power plant.

(14)
<u>去年１２月、奄美大島沖合いの東シナ海で起きた不審船事件で、沈没している船体をつり上げる</u>**クレーン台船が**、現場海域に到着し、明日引き揚げ作業が行われます。- the main event, the arrival of a crane vessel, is linked to a larger context of a suspicious ship incident that took place the previous year.

Explanatory grounding
(15)
<u>インターネット上で、映画や音楽などを不特定多数の人が手に入れられる、Winnyと呼ばれる、ファイル交換ソフトを開発した、東京大学大学院の助手</u>、**カネコ・イサム容疑者が**、著作権法違反の幇助の疑いで逮捕されました。- before introducing the main event, an arrest, the lead offers background information, the details of the misdeed that led to the arrest.
(16)
<u>今日午後、横浜市の金沢区で</u>、下校途中の小学生の列に**乗用車が**突っ込み、1人が死亡、4人が重軽傷を負いました。- information on time and place of the main event is given before actually introducing it.

      The frequency of these types of grounding in correlation to the types of subject NP referents in television and newspaper news leads is shown in Figures (4) and (5) below. This time, only first subject NP referents, introduced in the first main clause of the lead paragraph have been taken into account.

Fig. (4) Grounding and subject NP referent identifiability in Japanese television news leads

Types of referents	Frequency in the first lead sentence						Total	
	zero grounding		explanatory grounding		connective grounding			
	wa	*ga*	*wa*	*ga*	*wa*	*ga*	*wa*	*ga*
A-1. active	13	0	4	2	17	1	34	3
A-2. semi-active	7	3	19	12	14	1	40	16

B-3. identifiable – prev. introd.	—	—	0	2	2	1	2	3
B-4. identifiable – inferable	—	—	0	6	0	2	0	8
B-5. new, grounded	—	—	7	133	0	15	7	148
Total	20 (24.09 %)	3 (1.68 %)	30 (36.14 %)	155 (87.07 %)	33 (39.75 %)	20 (11.23%)	83 (100 %)	178 (100 %)

From these results, it appears that most subject NP referents, either topics or non-topics, tend to be introduced into the discourse accompanied by some kind of grounding information. Only 24.09% of the topics are directly introduced in the beginning of the lead sentence, with zero grounding. In 36.14% of the cases, topics are given an explanatory grounding, but more often (39.75%, especially in the case of easily identifiable, active) the accompanying grounding information does not bring additional explanations, but rather connects them to some other known events.

Non-topic subject NPs, on the other hand, which, in the majority of cases, have new referents of the B-5 type, tend to be accompanied by explanatory information (87.07%). Instances of connective grounding are less frequent (11.23%), and of zero grounding (such as example 9), very few (1.68%).

Fig. (5) Grounding and subject NP referent identifiability in Japanese newspaper news leads

Types of referents	Frequency in the first lead sentence						Total	
	zero grounding		explanatory grounding		connective grounding			
	wa	*ga*	*wa*	*ga*	*wa*	*ga*	*wa*	*ga*
A-1. active	6	0	2	0	3	0	11	0
A-2. semi-active	6	1	1	1	3	0	10	2
B-3. identifiable – prev. introd.	—	—	0	0	0	0	0	0
B-4. identifiable – inferable	—	—	0	0	0	0	0	0
B-5. new,	—	—	1	23	0	0	1	2

grounded								3
Total	12 (54.54%)	1 (4.00%)	4 (18.18%)	24 (96.00%)	6 (27.27%)	0 (0.00%)	25 (100%)	

Though based on a smaller number of occurrences, the results shown by the analysis of newspaper news suggest characteristics that are similar to the television news: topics tend to have connective grounding (27.27%), while the great majority of non-topic subject NPs have explanatory grounding (96.00%). In newspaper news we find, however, more instances of topics with zero grounding (54.54%)[6].

CONCLUSIONS

Subject NP referents that appear in the Japanese news leads are of two large types, (A) referents assumed to be *at least familiar* to the receiver at the moment they are introduced into the discourse, such as widely known national and international political, economical, social institutions or their representative individuals, and (B) referents assumed to be *unknown* to the receiver at the moment they are introduced into the discourse, such as individuals, things or particular events that have just happened.

A-type referents of subject NPs are usually introduced into the discourse directly as topics (marked with *wa*), while B-type referents are presented as non-topics, marked with the typical nominative case particle *ga*. Also, in order to make them acquire the *identifiable* status and help the receiver easily integrate them into his cognitive model, B-type referents are always accompanied by some kind of additional information (*grounding*).

Grounding information, before the first subject NP is introduced, is found in 98% of the cases where the subject NP is marked with *ga*, and even in 26% of the cases of topics (NP-*wa*). Grounding information tends to have different functions according to the type of NP referents. In most cases, topic subject NP referents have *connective grounding*, i.e. information that does not bring additional explanations on the referents, but rather connects them or the entire main event presented by the news item to some other events that are assumed to be known by the receiver. Non-topic subject NP referents, on the other hand, which, in the majority of cases, have new referents of the B-5 type, have *explanatory grounding*, i.e. additional information that helps the receiver *identify* them. Overall, both types of grounding, and the fact that grounding is provided *before* the introduction of

---

[6] However, further analysis of a larger newspaper corpus is needed in order to clearly establish whether zero grounding is characteristic to written news, or to specific types of news contents (such as politics, economy etc), or if these results are simply due to the relatively small number of instances analyzed.

new information, play an important role in information processing, making the new information easier to integrate into the receiver's preexistent cognitive model.

REFERENCES

1. BELL, Allan (1991) *The Language of News Media*, Oxford UK / Cambridge MA, Blackwell.

2. LAMBRECHT, Knud (1994) *Information Structure and Sentence Form*, Cambridge, Cambridge University Press.

3. LANGSTON, Mark; TRABASSO, Tom (1999) "Modeling Causal Integration and Availability of Information during Comprehension of Narrative Texts", in Herre van Oostendorp and Susan R. Goldman (eds.) *The Construction of Mental Representations during Reading*, Mahwah NJ / London, Laurence Erlbaum Associates, pp 29-69.

4. PRINCE, Ellen F. (1981) "Toward a Taxonomy of Given-New Information", in Peter Cole (ed.) *Radical Pragmatics*, New York, Academic Press, pp. 223-255.

5. SION, Andreea (2003)「テレビ・ニュースのリード文における有題文・無題文の選択」『筑波応用言語学研究』10, pp. 69-82.

6. VAN DIJK, TEUN A.; KINTSCH, Walter (1983) *Strategies of Discourse Comprehension*, New York, Academic Press.