

Nominalised CPs in Persian: a Parametric Account

Peter Öhl & Ahmad R. Lotfi

University of Munich / Esfahan Azad University

Persian argument clauses may be determined by the definite determiner *in* preceding the complementiser, and exactly like nominal objects, object sentences may be marked by the case marker *rā* occurring with specific direct objects. The Persian data strongly suggest that preverbal clauses are in fact DPs. We propose that the distribution of subordinate clauses is universally correlated with the ability to bear case. CPs in case position are universally selected by D° for the sake of the realisation of abstract case. Complementisers like Persian *ke* are in fact nominal heads selecting full clauses. Only languages with nominal complementisers can have full clauses in case positions.

Introduction

This paper mainly deals with the following two constructions in Standard Persian (*Fārsi*):

- (1) a. [_{DP} **In** [_{CP} **ke** [_{IP} gorbehā šir dust dārand]]] tabi'i ast.
DET – *that* – *cats* – *milk* – *like* – *have* – *natural* – *is*
'That cats like milk is natural.'
- b. Man [_{PP} [_{DP} **in** [_{CP} **ke** [_{IP} gorbehā šir dust dārand]]] **rā**] midānam.
I – DET – *that* – *cats* – *milk* – *like* – *have* – OBJ – *know*
'I know that cats like milk'

The examples show that in Persian, argument clauses may be determined by the definite determiner *in*¹ preceding the subordination marker (henceforth 'SUB'), and exactly like nominal objects, object clauses may be marked by the particle *rā*. According to Lotfi (1998) and Karimi (2003: 91), *rā* is a case marker occurring with specific direct objects.

¹ In section 2, we will extensively argue for the status of *in* as a determiner. In more traditional accounts, *in* is considered a demonstrative pronoun that is the syntactic object of the main clause predicate, whereas the CP introduced by *ke* has an attributive function. Compare:

(i) I know that, that cats like milk.

- (2) a. Man **in** doxtar **rā** didam.
I – DET – girl – OBJ – saw
 'I saw this girl.'
- b. **In** pesar to **rā** did.
DET – boy – you – OBJ – saw.
 'This boy saw you.'

We intend to explain the Persian data with reference to the comparative syntax of CP-embedding. There are many languages characterised as SOV where CPs may occur (and do so predominantly) in a postverbal position. Well known examples are Dutch, German, Hindi, Bengali and Persian.

- (3) a. Ich habe die Antwort (ACC) **gewusst**. (German)
I – have – the – answer – known
 'I knew the answer.'
- b. Ich habe **gewusst**, [_{CP} dass [_{IP} Katzen Milch mögen]]. (German)
I – have – known – that – cats – milk – like
 'I knew that cats like milk.'
- (4) a. Man in doxtar **rā** **didam**. (= 2a) (Persian)
I – DET – girl – OBJ – saw
 'I saw this girl.'
- b. Man **midānam** [_{CP} ke [_{IP} gorbehā šir dust dārand]]. (≠ 1b) (Persian)
I – know – that – cats – milk – like – have
 'I know that cats like milk.'

On the other hand, other SOV languages like Japanese or Korean do not exhibit such an asymmetry. CP objects precede the verb the same way nominals do. Moreover, DPs and CPs are overtly case marked by the same particle. Japanese, for instance, uses the postposition *o*.

- (5) Mary wa [_{PP} [_{CP} [John ga koohii **o** nomu] **no**] **o**] mi-ta
Mary – TOP – John – NOM – coffee – ACC – drinks – SUB – ACC – seePST
 'Mary saw that John drank coffee'.

It is not an accident that the Japanese SUB **no** has often been analysed as a nominalising particle (cf. Kaiser & al. 2001: 69). If converted by **no**, full clauses behave like nominal arguments with respect to case marking and syntactic distribution.

Persian allows preverbal CP objects optionally. It is exactly in this case, however, that it is obligatory to use **in** as a determiner. Moreover, it is crucial to use the direct object marker **rā** with CPs, which is also used with DP objects (cf. 2a/ 4a). We analyse this particle as a case postposition comparable to Japanese **o** above.

- (6) a. Man [_{PP} [_{DP} **in** [_{CP} ke [_{IP} gorbehā šir dust dārand]]] **rā**] midānam. (= 1b)
I – this – that – cats – milk – like – have – ACC – know
 'I know that cats like milk.'
- b. *Man [_{CP} ke [_{IP} gorbehā šir dust dārand]] midānam.
- c. *Man [_{DP} **in** [_{CP} ke [_{IP} gorbehā šir dust dārand]]] midānam.

The Persian data strongly suggest that preverbal clauses are in fact DPs. Since 'bare' CPs are ungrammatical in the canonical object position of DPs (6b), we must assume that the distribution of arguments is constrained by their category. Moreover, since the data also suggest that preverbal A-positions must not be occupied by elements not bearing abstract case (6c), which is overtly realised by a postposition in Persian, this categorial restriction seems to follow from the inability of non-nominals to realise case.

What is very striking is that the Persian CMP **ke** is optional if (and only if) the CP follows the verb. It is obligatory, if the CP precedes it (6).

- (7) Man **midānam** [_{CP} (ke) [_{IP} gorbehā šir dust dārand]]. (cf. 4b)
I – know – that – cats – milk – like – have
 'I know that cats like milk.'

Parallel to that, languages like German allow V2 clauses as objects only in postverbal position. An object clause must have a CMP if it is fronted.

- (8) a. Jeder weiß, [_{CP} Katzen mögen Milch].
everybody – knows – cats – like – milk
- b. *[_{CP} Katzen mögen Milch], weiß jeder.

- c. [_{CP} **Dass** Katzen Milch mögen], weiß jeder.
that – cats – milk – like – everybody – knows

Our proposal is that the distribution of subordinate clauses universally correlates to the category immediately dominating them. We also propose that the distribution of subordinate clauses is universally correlated with the ability to bear case, and that CPs in their canonical argument position – i.e. a case position – must be DPs. CPs in case position are universally selected by D° for the sake of the realisation of abstract case. SUBs like Persian *ke* are in fact nominal heads selecting full clauses. Only languages with nominal complementisers can have full clauses in case positions.² Languages that don't must move them to a derived non-argument position, which is the postverbal position where we find argument clauses in languages like German and Persian.

Determiners, Case Marking and CPs in Modern Persian

In modern Prs., there are two DETs that are used with both nominals and with clauses: *īn* and *ān*. They do not carry ϕ -features. If they are used with nouns, their distribution roughly compares to that of *this* and *that* in English.

- (9) To *īn* pesar-*o* didi, man *ān* doxtar-*o* didam.
you –this – boy-ACC – saw – I – that – girl – ACC – saw

'You saw this boy and I saw that girl.'

Therefore, these DETs have been classified as demonstratives in traditional grammar. Since they can be used as demonstrative pronouns, too, preverbal CPs are often regarded as attribute clauses that are adjoined to a NP headed by *īn*. However, Persian *wh*-object-clauses show that this cannot be the whole picture. Like internal *that*-clauses, *wh*-argument clauses preceding the finite verb are determined and case marked. The major differences are, first, that with *wh*-clauses, *ān* is used as the one specific DET. The use of the DET *īn* renders the sentence ungrammatical, which strongly points to the subcategorisation of *ān* for the selection of a *wh*-feature.

² Following Chomsky (1995) in assuming that only exceptions from primitive properties are listed in the lexicon, it seems that *c-selection* can thus be removed from the lexical entry of predicates. If we, based on case assignment, generalise the category of structural arguments as nominal (cf. also Pesetsky 1982), PP-objects do not fall under *c-selection* either: prepositions are specific realisations of abstract case.

(10) *Man **in** če (ke) to maxfi mikoni **rā** midānam.

I – DET – what – SUB – you – hide – do – ACC – know

'I know what you're hiding'

Second, that the ACC-marker **rā** can be attached to either the wh-word (if it is fronted), or to the whole clause.

(11) a. Man **ān** če ke to maxfi mikoni **rā** midānam.

I – DET – what – SUB – you – hide – do – ACC – know

b. Man **ān** če **rā** ke to maxfi mikoni midānam.

'I know what you're hiding'

This asymmetry can be explained by two basic assumptions: First, **ān** in (11) is a DET specified for a *wh*-feature. Second, it determines either a *wh*-clause (11a) or a *wh*-NP (11b). Whereas in (a), there is a *wh*-object-clause with an (optionally) doubly filled CP, the *wh*-NP is the syntactic object of the predicate *know* and has a **ke**-clause as a modifier in (b). In this case, the **ke**-clause can also be extraposed. In an unmarked clause with this reading, however, the *wh*-word **če** would be replaced by the full noun **čīs** (+ indefiniteness marker **-i**).

(12) a. Man **ān** če **rā** midānam ke to maxfi mikoni.

b. Man **ān** čīs-i **rā** midānam ke to maxfi mikoni.

Note that *wh*-elements are also case marked by **ra**, if they are the direct object of the embedded clause. If a *wh*-element is fronted, the case particle is optional. Thus, it is possible (though stylistically marked) to attach the case particle to both the moved *wh*-word (object of the SubC-verb) and the whole *wh*-clause (object of the MC-verb).

(13) man **ān** [che, **rā** [_{C'} [_{IP} to t_i maxfi mikoni]]] **rā** midānam

This is direct evidence for our assumption that a whole CP, and not only a pronoun with a CP-attribute, is the syntactic object of the MC-predicate. The strongest argument, however, is the mere fact that there are specific DETs for *wh*- and non-*wh*-clauses³. If we assume that **īn** and **ān** are definite determiners specific to

³ Note that *wh*-fronting in Persian is optional. If and only if the *wh*-word is fronted, there is the option to drop the CMP **ke**. In *wh*-in-situ-clauses, **ke** is obligatory, the DET **in** is used and the case particle follows the clause.

argument clauses, both their complementary distribution and their obligatory presence are self-explanatory. It is implausible to assume that a *wh*-attribute-clause is used only with a specific demonstrative pronoun. That *ān* occurs all and only with *wh*-clauses can only be explained through a selectional restriction: the DET *ān* is subcategorised for a *wh*-complement, whereas *īn* is subcategorised for a non-*wh*-complement. Because argument clauses are always definite⁴ and definiteness of nominals must be indicated in Prs., clauses in argument position are obligatorily determined by either *īn* or *ān*. Persian argument clauses are constructed like nominal arguments, if they are in their canonical argument position.

Our assumption is that the Persian CMP *ke* carries a nominal feature giving a clause the ability to receive structural case. However, the feature [N] can optionally be represented, or lexicalised, by a fronted *wh*-element. In the following paragraphs, we derive these properties from the different grammaticalisation process producing the grammatical markers *in/ān* and *rā*, and especially *ke*. *ke* has developed from a relative marker. *In* and *ān*, however, have been grammaticalised as definite determiners specific to one of the two kinds of CPs. Together with *ke*, they yield the possibility of overt case marking, since only DPs can be assigned case.

Contrastive Data on Subordination

The Feature Structure of the C-Domain

We assume that, besides features denoting semantic markedness like clause mood, FOC etc., the following formal features can be realised in the outmost shells of argument clauses: CASE, DEF, N, *wh*. Parametric change effects the way these features are realised. Some variation can be shown in German, English and Dutch.

- (14) a. Ich frage mich [_{CP} *wer* [_{IP} so etwas liest]] (standard German)
 I – know – not – who – such – thing – reads
- b. I frag me [_{CP} *wer* [_C *dass* [_{IP} so äps liast]]] (Swabian)

-
- (i) Man *in* ke to če rā maxfi mikoni *rā* midoonam.
 I – DET – CMP – you – what – ACC – hide – do – ACC – know

Thus, *ān* seems in fact to be subcategorised for a strong *wh*-feature.

⁴ Clauses refer to propositions, which are always definite.

I – know – not – who – that – such – thing – reads

'I wonder who would read such stuff.'

The German standard sentence (a) shows that there is no CMP lexicalised together with a fronted *wh*-word. In some varieties, however, the CP is doubly filled in a sentence with exactly the same interpretation, and it seems plausible that there are two words lexicalising (at least) two features that are lexicalised by one word in the standard. Examples from colloquial Dutch are even more striking:

- (15) Ik vraag me af [_{CP} **wie** [_? **of** [_? **dat** [_{IP} taalkunde studeert]]]]
I – ask – me – of – who – if – CMP – linguistics – studies

I wonder who studies linguistics.

This shows that there are at least three relevant features to be lexicalised. These are the feature related to *wh*-focus, the feature of propositional disjunction specific to interrogatives (cf. Bayer 2004, 66; further on symbolised by 'Q') lexicalised by **of**, and a feature of syntactic subordination lexicalised by **dat**. The distributional facts must be a matter of parameterisation, which is confirmed by a Middle English example, where doubly filled CPs were – in contrast to Modern English – still possible.

- (16) men shal wel knowe **who that** I am. (Caxton R67; ~1485 AD; in
'Men shall well know who I am.' Lightfott 1979, 322)

In this case we suggest that both *wh* and *Q* are phonologically represented by the *wh*-word, whereas SUB is lexicalised by **that**.

Categorical Status and Distribution of Argument Clauses

Whereas the occurrence of several items in CP has a suggestive explanation in the presence of different heads representing different features, another phenomenon is a much bigger puzzle, which we want to demonstrate at the distribution of object clauses in German. Like Persian, German has an SOV-base. The first fact to observe is that V2-clauses, i.e. clauses without CMP, cannot occur sentence internally – they must be extraposed, i.e. follow the verb.

- (17) a. Ich habe niemals behauptet, [das stimmt alles].
I – have – never – claimed – this – is-true – all

'I never claimed all of this is true.'

- b. *Ich habe [das stimmt alles] niemals behauptet.
I – have – this – is-true – all – never – claimed
- c. *[das stimmt alles] habe ich niemals behauptet.
this – is-true – all – have – I – never – claimed

that-clauses may occur in several positions:

- (18) a. Ich habe [_{CP} dass das alles stimmt] niemals behauptet.
I – have – that – this – all – is-true – never – claimed
- b. [_{CP} Dass das alles stimmt] habe ich niemals behauptet.

Second, the German data strongly suggest that dropped CMPs are not only PF-Ø. Embedded clauses lacking a CMP are always V2. The verb in second position and the CMP are in complementary distribution.

- (19) *Ich habe niemals behauptet, [dass das stimmt alles].

Thus, there is an obvious correlation between the distribution of argument clauses and the presence of a CMP. The dropped CMP does not tell the whole story, however. V2-complements in German are structured exactly like main clauses. In fact, we would like to suggest that their categorial status is different from that of subjects and objects introduced by a CMP. In German, V2 complements are very clearly restricted to *epistemic* and *assertive* verbs (cf. Öhl 2003, 165; Meinunger 2004, 315f ; for a list see Vikner 1995, 71). The context licensing CMP-drop in Gm. also licenses topicalisation or focus fronting of constituents in the embedded clause.⁵

- (20) a. sie sagen, der Junge studiert Linguistik
they – say – DET – boy – linguistics – studies
- b. ich behaupte, am besten studiert der Junge Linguistik
I – claim – at – best – studies – DET – boy – linguistics
'I claim it will be the best for the boy to study linguistics.'
- c. ich glaube, Linguistik studiert der Junge
I – believe – linguistics – studies – DET – boy

Verbs of other semantic classes (e.g. *factives*, *volitionals*) do not license constituent fronting in embedded clauses, and thus neither V2 nor CMP drop.

⁵ A similar observation has been made by Poletto (2000: 123ff) for Italian.

- (21) a. *sie leugnen, Linguistik studiert der Junge
they – deny – linguistics – studies – DET – boy
- b. *ich bezweifle, Linguistik studiert der Junge
I – doubt-on – linguistics – studies – DET – boy
- c. *ich bin überrascht, Linguistik studiert der Junge
I – am – amazed – linguistics – studies – DET – boy
- d. *ich erwarte, Linguistik studiert der Junge
I – expect – linguistics – studies – DET – boy
- e. *ich will, Linguistik studiert der Junge
I – want – linguistics – studies – DET – boy
- f. *ich befehle, Linguistik studiert der Junge
I – order – linguistics – studies – DET – boy

The data above suggest, that V2 complements in fact represent embedded *assertive speech acts* (cf. also Meinunger 2004: 315f). Indeed data from several languages suggest that idiosyncratic constraints on embedding speech acts may be the reason for restrictions on the selectional frame. In colloquial Norwegian, there is a stylistically marked option to drop the CMP in SubCs embedded by exactly the V allowing embedded V2 in Gm. Those V also license embedded focus fronting (cf. Öhl 2003: 167ff).

- (22) a. jeg hevder (at) gutt-en studerer lingvistikk
I – claim – CMP – boy-DET – studies – linguistics
- b. jeg vet (at) gutt-en studerer lingvistikk
I – know – CMP – boy-DET – studies – linguistics
- c. jeg tror (at) gutt-en studerer lingvistikk
I – believe – CMP – boy-DET – studies – linguistics
- d. de sier (at) gutt-en studerer lingvistikk
they – say – CMP – boy-DET – studies – linguistics
- (23) a. jeg hevder (at) lingvistikk studerer gutt-en
I – claim – CMP – linguistics – studies – boy-DET
- b. jeg vet (at) lingvistikk studerer gutt-en
I – know – CMP – linguistics – studies – boy-DET
- c. jeg tror (at) lingvistikk studerer gutt-en
I – believe – CMP – linguistics – studies – boy-DET

- d. de sier (at) lingvistikk studerer gutt-en
they – say – CMP – linguistics – studies – boy-DET

Vs not licensing CMP drop in colloquial Nw. do not license embedded V2 either.

- (24) a. jeg forventer *(at) gutt-en studerer lingvistikk
I – expect – CMP – boy-DET – studies – linguistics
- b. jeg er overrasket over *(at) gutt-en studerer lingvistikk
I – am – surprised – about – CMP – boy-DET – studies – linguistics
- c. de benekter *(at) gutt-en studerer lingvistikk
they – deny – CMP – boy-DET – studies – linguistics
- d. jeg tviler på *(at) gutt-en studerer lingvistikk
I – doubt – on – CMP – boy-DET – studies – linguistics
- (25) a. *jeg forventer at lingvistikk studerer gutten
- b. *jeg er overrasket over at lingvistikk studerer gutten
- c. *de benekter at lingvistikk studerer gutten
- d. *jeg tviler på at lingvistikk studerer gutten

More evidence for our assumptions comes from Russian (cf. Öhl 2003: 167). The language has a focus-phrase in the C-domain below the layer hosting the CMP exactly with the verbs in question, and exactly with these, the CMP can be dropped.

- (26) a. Ja dumaju, [_{CP} (čto) [_{FocP} lingvistiku [_{IP} mal'čik [_{I'} budet [_{VP} izučat']···]]
*I – think – CMP – linguistics – boy – will – study*INF
- b. Ja dumaju, [_{CP} (čto) [_{FocP} lingvistiku [_{IP} v Valentinov den' nikto [_{VP} s'udovol'stvijem učit' ne zaxočet]···]]
I – think – CMP – linguistics – on – Valentine's – day – nobody – gladly – learn – NEG – want(FUT)-(PRF)

On the other hand, factive and emotive verbs like the following ones allow neither CMP drop nor embedded focus fronting:

- (27) a. Ja otricaju, *(čto) mal'čik izučajet lingvistiku.
I – deny – that – boy – studies – linguistics
- b. Ja ožidaju, *(čto) mal'čik budet izučat' lingvistiku.
I – expect – that – boy – will – study – linguistics

- c. Ja byl udivljon, *(čto) mal'čik izučajet lingvistiku.
I – was – surprised – that – boy – studies – linguistics
- d. Ja somnevajus', *(čto) mal'čik izučajet lingvistiku.
I – doubt – that – boy – studies – linguistics

The question of why assertive verbs in these languages license embedded topic or focus has a simple and intuitive answer. They obviously allow the clauses they are selecting to autonomously project features of information structure (Top, Foc), which are usually administrated by illocutionary force. Since they license embedded assertions, the subordinate clause has autonomous organisation of information that can be structured like in main clauses.⁶

Case and Clauses

Arguments and Case Assignment

In certain languages, embedded assertions do not have to be selected as arguments, but may be adjoined in an extraposed position coindexed with the argument position that may be occupied by an empty pronominal (*pro*). Complements generated in their argument position, however, require a CMP as the clause-specific representation of the feature [N]. This feature is the precondition for case assignment.

Thus, there are three logical options for structural complements of verbs selecting clauses:

1. CPs as complements of non-assertive Verbs are always in an argument position, do not project as TopP nor FocP, and must be selected by a nominal head.
2. CPs as complements of assertive Verbs project as TopP or FocP. They are also selected by a nominal head when occurring in an argument position.
3. CPs that are complements of assertive Verbs but are not in an argument position project as TopP or FocP and may occur without a nominal head.

⁶ A similar point has been made by van Gelderen (2003), who shows that in English and Dutch only the clausal complements of assertive verbs project TopP and FocP.

Embedded assertions have an extraordinary status with respect to the extended projection of C. Moreover, they are not subject to the case principle and therefore may lack a CMP, if they are generated outside of VP. The following basic assumptions are crucial for our analysis:

1. Positions to the right of V in OV-languages are no argument positions but rather adjunct positions. CMP-drop-clauses are generated in such a position universally.
2. Verbs c-select nominals. In order to be licensed as an argument, a syntactic element must carry a nominal feature. In the case of clauses, this nominal feature can be lexicalised by a subordination marker.
3. Syntactic elements are assigned case in specific functional positions – say AgrS, AgrO etc. If and only if a language is parametrically able to realise the case of clauses, it can have CPs in an argument position.

We assume that there is a universal constraint on case assignment to all arguments in their canonical argument-position. *Abstract case* expresses a specific relation between a head X° and its argument(s). This structural relation can interact with conceptual features (i.e. *lexical case* realising a semantic role, cf. Culicover 1997, 44) or be purely syntactic (identifying a structural relation, e.g. the NOM – ACC alternation under passivisation). Since arguments are thought to be assigned abstract case by V or Agr (cf. Chomsky & Lasnik 1995), the need for case has been phrased as a well known syntactic principle.

(28) **Case Principle** (cf. Chomsky & Lasnik 1995: 561)

Every realised DP/ NP must be assigned abstract case. A chain is visible for Θ -marking if it contains a case-position.

From the second statement above it follows that Θ -marking of CPs must be made visible by case. If we assume that case is assigned to XPs by V or Agr, the straightforward question arises as to what happens if the argument is a clause. Agr is a structural relation between V° and its arguments, or more formally, the correspondence between a structure of ϕ -Fs associated with case on the nominal side, and the same ϕ -features on the verbal side. There is no reason why nominal and clausal arguments should be subject to different case conditions only because they lack overt case marking in many languages. Since it is implausible that only

CP-arguments should be exempt from this principle, we adopt the case principle in a generalised version.

(29) **Case Principle** (amended)

Every argument chain must be licensed by abstract case.

Case assignment is a constraint on *formal identification* in syntax. Predicates assigning Θ -roles select nominals since the formal representation of their arguments must be identified by case. This we also take to be true for the selection of subject and complement clauses. In order to be able to be licensed by case, they need a grammatical marker turning them into a nominal category.

We propose that the role of SUBs is to make clausal arguments able to be in a chain with a case position. It was suggested already by Kayne (1984) that C was a *potentially nominal category* (cf. Müller & Sternefeld 1990: 37ff)⁷. Müller (1993:60) found some evidence from data on extraction from embedded clauses, that CPs are in fact NPs. He therefore formulated the *NP-shell Hypothesis*.

(30) **The NP-Shell Hypothesis**

All embedded finite clauses are NPs with a phonetically empty head.

Alternatively, we assume that CMPs themselves are heads carrying the categorial F [N]. They derive a nominal category from a verbal one on the level of syntax. A CMP must be present since it represents a nominal F assigned to the clause in order to make it visible for case marking. This means, however, that complement clauses without CMPs are in fact not selected as an argument of V. Some consequences are reflected by the phenomenon that has been called *CMP-drop* in the literature. There is evidence for varying restrictions on extraordinary licensing of complement clauses in many languages. Some languages never allow CMP-drop, like Latvian, French (cf. Öhl 2003: 159f) or Greek (Öhl 2003: 200). Languages like English seem to be quite liberal, but not even in English, however, all verbs license complement clauses without a CMP.

(31) *John regrets/ expects/ understands Mary studies linguistics.

⁷ According to Müller & Sternefeld (1990), this is the reason why *Wh*-Fs are generated in SPEC/C, whereas Topics are in the SPEC of the verbal extension that we call FocP, cf above, Ch. 1.2.5.

This shows that the so called 'CMP-drop-phenomenon' is not a mere configurational problem, as it was suggested by Stowell (1981). He proposed that CMPs may be dropped if they are in a governed position (which was repeated in a minimalist framework by Pesetsky & Torrego 2007). In contrast, we suggest that CMPs can be dropped, if they are subject to specific licensing conditions that are determined by the embedding verb. In our terms, this is explained by a parameter for licensing logical complements of specific verbs in a syntactic adjunct position, which therefore do not fall under the case principle. Persian is a language that provides direct evidence for the conditions on the realisation of case that are assumed in this paper.

CASE and DET above CP: Analysing the Persian Evidence

The Persian transitive verb does not require (preverbal) object position to be filled in by a nominal category, but it requires overt representation of abstract case. If the object position of the predicate is empty, the verb will be attached an object-morpheme (e.g. the third person singular clitic object **-esh**). In this case we assume that an object *pro* holds the position.

- (32) a. Man **ān** rā didam.
 I – him/her – saw
- b. Man **pro** didam-**esh**.
 I – pro – saw-him/her
- 'I saw him/her.'

The distribution of Persian argument clauses meets the generalisations made above about assertive verbs in the languages discussed above: The object clause may follow the verb, and in this case the CMP **ke** is optional.

- (33) a. (Man) fekr ne-mikardam [(**ke**) pesar-e zabanšenāsi bexānad]
 I - thought - didn't - CMP – boy-DEF- linguistics – studies
- 'I did not think that the boy would study linguistics.'
- b. (Man) midānestam [(**ke**) pesar-e zabanšenāsi xāhad xānd]
 I - knew - CMP - boy - linguistics - FUT - studied
- 'I knew that the boy would study linguistics.'

This is also true if the extraposed clause is a *wh-in-situ*-interrogative.

- (34) Man midānam [(*ke*) to če *rā* maxfi mikoni]
I – know – you – what – ACC – hide – do

There are three basic distributional options.

- (35) a. Man *pro* midānam (*ke*) zamin gerd-e.
I – know-1stsg-pres – CMP – the – earth – round-is
- b. Man *in* *rā* midoonam *ke* zamin gerd-e.
I – this – ACC – know – CMP – the – earth – round-is
- c. Man *in* *ke* zamin gerd-e *rā* midānam.
I – DET – CMP – the – earth – round-is – ACC – know-1stsg-pres
- 'I know that the earth is round.'

We take these structures as three logically equivalent projections, not as derivations from one base projection. For this, we put forward an argument from derivational economy, given Chomsky's (1995 etc.) stipulation that overt movement operations are costly. If *pro* checks the features in AgrO, sentence (a) can pass LF, nothing forcing the CP to move to the object position. Second, since the insertion of further lexical elements is necessary if the CP precedes V, a movement process like this should be excluded by representational economy, too. We take this as an indicator that there is the possibility to logically define a (phonologically zero) syntactic object by a postposed proposition that might be right adjoined to the VP or higher. The main difference between (a) and (b) is that there is a demonstrative pronoun in the object position in (b) functioning as an overt correlative. We argue that there are no logical but only performance based differences, making use of the deictic force of a demonstrative. The striking restriction with respect to our claims is that in (b), the extraposed clause must not drop the CMP. The reason must be that *in* is a nominal head co-referential with a nominal CP rather than a non-nominal S. In (c), however, the clause is in object position.

If there are three ways of base generation, there are three equally economical structures and there is no need to explain optional movement. To derive one or the other structure from (c) above would make additional stipulations necessary, e.g. why a CMP should be optionally inserted before, but obligatorily after movement, or why *in* and *rā* never occur with a postverbal CP. Thus, there is a regular representation where object clauses occur in extraposed position to the left of V. Once it is semantically and syntactically licensed, a CP headed by the syntactic

subordinator **ke** or a bare S (i.e. a CP without the head hosting the feature complex SUB+N) can be inserted in the post-verbal position, with the object *pro* in object position.

Only extraposed clauses that are not in the canonical Θ -position may lack the CMP. However, as soon as a CP is in Θ -position (i.e. the subject or a preverbal object), the CMP is obligatory. Since subjects are never extraposed in Persian, they never drop the CMP. And as soon as one condition for the obligatoriness of **ke** is given, also the DET **in** must precede the clause. Moreover, the object clause must have the postponed marker of ACC case **ra**.

- (36) a. Man midānam ([_{CP} ke] [_{VP} gorbehā šir dust dārand]). (SV - object-CP)
I – know – CMP – cats – milk – like – have

'I know that cats like milk.'

- b. [_{DP} **In** [_{CP} ke [_{VP} gorbehā šir dust dārand]]] tabi'i-ye. (SV)
this – that – cats – milk – like – have – natural – is

'That cats like milk is natural.'

- c. Man [_{DP} **in** [_{CP} ke [_{VP} gorbehā šir dust dārand]] **rā**] midānam. (SOV)
I – this – CMP – cats – milk – like – have – ACC – know

'I know that cats like milk'

The Persian data not only indicate that clauses generated in argument position must have an overt CMP, but also that CPs are determined and overtly case marked. This means as soon as an argument clause is not extraposed, it behaves like an NP rather than a sentence. It is a DP with a CP complement (cf. Lotfi 1998; Öhl 2003: 185), i.e. a clause with a nominal head.

Prs. object clauses must be determined and provided with the postpositional accusative PTC **ra**, if they precede V. Before the ACC marker and DET can apply, however, the proposition must be nominalised by CMP. The CMP is very clearly crucial for both case marking and the use of a determiner; it must be the element changing the category of S to [N]. Therefore, there are no CMPless object clauses preceding V. And there are no preverbal argument clauses that are not case marked or are not determined.

- (37) a. *Man [_{DP} **In** [_{VP} gorbeha shir doost darand] **rā**] midoonam.

- b. ***Ke** gorbeha shir doost darand tabi'i-ye.

This raises the question of how case is structurally represented. Note that in Persian, there are also CPs as PP-objects and PP-adverbials.

- (38) a. Man [_{PP} **vase** [_{DP} in [_{CP} ke [_{TP} autobus biyād]]] vā-istādam.

I – for – DET – CMP – bus – comes – wait

'I'm waiting for the bus to come.'

- b. Man [_{PP} **be** [_{DP} in [_{CP} ke u nayāmad]] e'terāz dāram

I – to – DET – that – s/he – not-came – objection – have

'I have an objection to him/her not showing up.'

- c. Man [_{PP} **az** [_{DP} in [_{CP} ke u nayāmad asabāni šodam]]

I – from – DET – that – s/he – not-came – angry – became

'I got angry as s/he didn't show up.'

- d. Man mimānam [_{PP} **tā** [_{DP} in [_{CP} ke [_{TP} to biyāyi]...]]

I – stay – till – DET – CMP – you – come

'I wait till you come.'

We suggest that the selection of nominals by prepositions and by case particles are parallel. Blake (2001: 165) considers the categorial status of **rā** as a *postposition* (cf. Öhl 2003: 185). As Don Stilo (p.c.) pointed out to us, it is even undesirable from the typological point of view to consider **rā** a postposition, since there aren't any lexical postpositions in Persian. It is well-known that in many languages, there are adpositional objects alternating with objects marked by structural case. These adpositions don't have any lexical meaning and seem to be nothing else but a lexical representation of abstract case. Thus, it is a natural assumption that case assigned by V and prepositions governed by V are structurally represented by the same kind of functional head, which we label KP (cf. Ghomeishi 1997 for the head K° hosting **rā**).

- (39) a. I am expecting [_{KP} ACC [_{DP} the bus]]

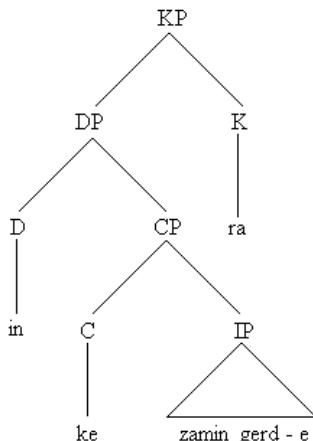
- b. I am waiting [_{KP} for [_{DP} the bus]]

In languages like English, structural accusative does not have a morphological representation. In Persian, it does. **Ra** realises a 'case-head', formally paralleling a postposition. We claim that both the selection of a syntactic element by P° and the

assignment of case presuppose the presence of a category D° , which in turn presupposes an element of the category $[+N]$. The feature $[+N]$ can be represented by a lexical head N° or the specific head in the C-Domain carrying the feature $[N]$. The structure of an argument clause in Persian can be drawn as follows:

(40)

(cf. 35c)



Alternatively, the Persian CP can be selected by a lexical preposition (cf. 38 above). The assumption that overt case assignment to CPs and selection by a prepositional head are parallel is confirmed by English data discussed by Pesetsky (1982). English has prepositional NP-objects, but it does not have prepositional CP objects.

(41) a. John asked the time.

b. John wondered *(about) the time.

(42) a. John asked what time it was.

b. John wondered (*about) what time it was.

We can make the following generalisation:

(43) ***Generalisation on Case Marking of Clauses***

Languages that do not have PP-clauses do not case mark CPs overtly.

If we assume that overt case assignment universally depends on the presence of a category D° , it does not come as a surprise that subject clauses in the canonical

position must be preceded by an overt determiner in several languages, as the following data taken from Öhl (2003, 186) show:

- (44) *(To) óti spudázi glossología, den vlápti. (Greek)
 DET – CMP – *studies* – *linguistics* – NEG – *harm*
 'That he studies linguistics does not harm.'
- (45) a. *(Tas) ka viņš deĵoja mums patika (Latvian)
 DET – CMP – *he* – *danced* – *us*DAT – *amused*
 'That he danced amused us.'
- b. *(Tas) ka viņš mēcās lingvistiku viņam nekaiš.
 DET – CMP – *he* – *studies* – *linguistics* – *him*DAT – NEG*harm*
 'That he studies linguistics does not harm him.'

Different positions for Complementisers, Case Assignment and Determination

In the Russian example in (26), it could be seen that the CMP is generated above FocP. This is in fact the general assumption about the layered CP since Rizzi's (1997) model, where CMPs are generally positioned in Force°.

- (46) Credo [_{ForceP} [_{Force} che [_{TopP*} domani, [_{FocP} questo, [_{TopP2*} a Gianni [_{FinP} [_{IP} gli
 dovremo dire] (Rizzi 1997:295)
thinkIstsg – CMP – *tomorrow* – *this* – *to* – *John* – *him*DAT – *mustIstpl-FUT* –
sayINF

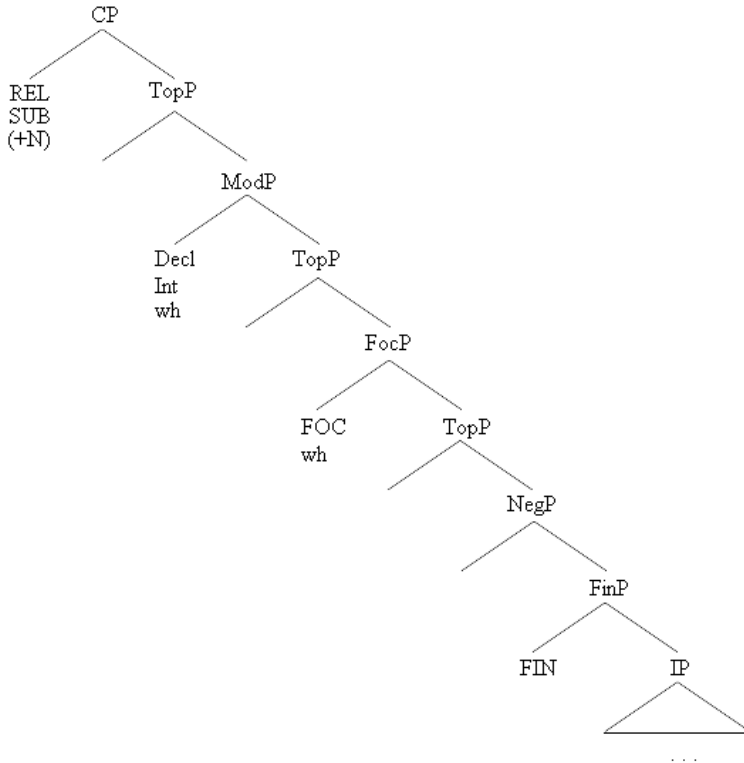
According to cross linguistic evidence from Italian, Greek and Persian, there are two different C-domains in matrix and embedded clauses, however:

(47) Syntactic Structure of the left periphery (C-Domain)

matrix: [_{ForceP} [hanging topic [scene setting adv. [left dislocation [list interpr.
 [contr.foc1 [contr.foc2 [inform. foc [_{FinP}] (Beninca&Poletto 2004: ex. 58)
 embedded: [_{CP} [... [_{ModP} [... [_{Neg} [_{FinP}] (Öhl & Korn 2006: 172; cf. Roussou
 2000: 79)

If we identify the formal feature of syntactic subordination (SUB) with the categorial feature [N], we end up with the following *potential* positions in the C-system more or less specified for the following kinds of features they can host (cf. Öhl & Korn 2006: 172; adapted from Rizzi 1997):

(48)



There is a number of implications to follow: there is rich empirical evidence that a lexical element phonologically representing SUB may be generated in a lower head. This is certainly true for the case where *wh*-elements lexicalise it, and is also suggested by the Dutch data, where the CMP is generated below the marker of clause mood.

(49) Ik vraag me af [_{CP} **wie** [_? **of** [_? **dat** [_{IP} taalkunde studeert]]]]
I – ask – me – of – who – if – CMP – linguistics – studies

There is also much evidence that German **dass** is generated in a lower CP layer (cf. Grewendorf 2002: 236f). We assume that there are (at least) two syntactic heads in the C-Domain that can be associated with a feature of syntactic subordination. The

lower head is also associated with features of the clause type.⁸ If a language generates a CMP there, it will be specified for the clause type, like the CMP of embedded interrogatives in German or English. If a language generates a CMP in the higher head, it will specify the clause type morphologically. In our view, this indicates that adding to the head Mod° indicating the clause type, there is an additional head C° projected only in subordinate clauses. It hosts a feature of syntactic subordination, or even just the categorial feature [+N]. This is why Persian has a particle indicating clause mood in addition to *ke* in interrogative object clauses.

- (50) a. Weisst Du, *ob* er Linguistik studiert?
 b. Do you know *if* he studies linguistics?
 c. ū porsīd [_{CP} *ke* [_{ModP} *āyā* [_{IP} man zabānšenāsī xānde būdam]]
he/she – asked – SUB – INT – I – linguistics – studied-had
 'He asked if I had studied linguistics.'

That Persian *ke* has a subordinating function that is much less specific than that of a complementiser like Gm. *dass* is also shown by its use in relative clauses and in different contexts of adverbial modification (cf. Korn & Öhl 2006: 140f).

- (51) kesī *ke* to dīde-ī emrūz raft. (Persian; Lazard 1992:229)
someone – SUB – you – have -seen – today – went
 'Someone that you saw went away today.'
- (52) a. nazdīk *ke* āmad ū-rā šenāxtam. (Lazard 1992:238)
 NP *near – SUB – came.3SG – s/he-FOC – recognised.1SG*
 'When s/he came near I recognised her/him.'
- b. raftam *ke* ān ketāb-rā bexaram. (Lazard 1992:218)
 NP *went.1SG – SUB – DEM – book-FOC – buy.1SG*
 'I went to buy the book.'

Moreover, it can be used as a quotative particle indicating quoted speech.

⁸ Note that Rizzi (2001) showed that the CMP *se* introducing embedded interrogatives in Italian is generated in a lower position than the SUB *che*. This is direct evidence from a single language that there are two positions for complementisers, the lower one associated with the clause type.

(53) Pesar goft (*ke*) man yek rūz zabānšenāsī xāham xānd.
boy – said.3SG – SUB – I – one – day – linguistics – will.1SG – study

'The boy said: One day, I will study linguistics.'

We assume that the head SUB is only specified for subordination.⁹ Particles generated there do not carry any additional information, and it may even depend on conventional rules how much underspecification of subordinate clauses a language allows. English has its *that* in relative and argument clauses, Italian has *che* in relative, argument and several adverbial clauses. Persian uses the subordinator even in quoted speech. German *dass*, however, is specified for declarative argument clauses. It is generated in the lower head indicating the clause type, or even in Fin°, as suggested by Grewendorf (2002: 236). This is the reason why, in contrast to the Scandinavian languages, V2 clauses in German and Dutch are in complementary distribution with clauses introduced by a CMP.

Indeed, German *dass*-clauses may be in preverbal position (in contrast to V2-complements). However, they must not be in the argument position where they are generated. Even though object clauses introduced by the CMP *dass* can occur to the left of the verb, they cannot be sister of it – unless they are placed at its right.

(54) a. *Ich habe niemals [_{CP} dass das alles stimmt] behauptet.

b. Ich habe niemals behauptet [_{CP} dass das alles stimmt] .

If complement clauses in German occur pre-verbally, they are either scrambled, in topic-position or in focus-position, which we take the pre-verbal position in (18a) to be. They are not, however, licensed in their argument position as the left sister of V (cf. 54a).

Persian has many distributional properties in common with German. The differences, however, can be explained through determination and case assignment. The data from the comparative syntax of Persian and German in our view suggest that only if the feature of syntactic subordination SUB is realised in the upmost CP-shell, it makes the categorial feature [N] visible for c-selection. Visibility of [N], however, is the *conditio sine qua non* for being selected by the functional head D°. D is the category that must be present prior to case assignment. And being

⁹ For a similar view cf. Klaiman (1976: 7).

assigned case is crucial for occurring in a canonical argument position. This is why German argument clauses must be generated in a non-canonical position – as they can in Persian. From there, they can move to higher positions like SPEC/TopP, but they cannot move to case positions where only DPs may be positioned. As in all cases of arguments not to be placed in case positions, case is assigned to a coindexed *pro*.

Conclusions

- Only case marked arguments can occur in a case position.
- Clauses may occur in a case position if a language parametrically has the option to case mark CPs.
- Persian data show that case marked CPs must be determined. It can be concluded that before the realisation of abstract case, a phrase must be selected by D°.
- D° selects phrases of the category [N]. It can be concluded that subordination markers like Persian *ke* and Japanese *no* have a categorial feature [N].
- Languages that do not have subordination markers of the category [N] cannot case mark full clauses, thus they cannot license full clauses in a case position, thus they must move them to an external (extraposed) position. This is the position where languages allow COMP-drop.

References

1. Bayer, Josef (2004). Decomposing the left periphery. Dialectal and cross-linguistic evidence. In: Horst Lohnstein & Susanne Trissler (Hgg.) (2004). *The Syntax and Semantics of the Left Periphery*. Berlin, New York: de Gruyter. 59-95.
2. Beninca, Paola & Cecilia Poletto (2004). Topic, Focus and V2. Defining the CP-Sublayers. In: Rizzi, Luigi (ed.). *The Structure of CP and IP. The Cartography of Syntactic Structures*. Vol. 2. Oxford University Press.
3. Blake, Barry J. (2001). *Case*. Cambridge: Cambridge University Press.
4. Chomsky, N. (1995). *The Minimalist Program*. Cambridge (MA): MIT Press.

5. Chomsky, N. & H. Lasnik (1995). Principles and Parameters in Syntactic Theory. In: Jacobs, J./Stechow, A.v./Sternefeld, W./ Vennemann, T. (eds) (1995). *Syntax: An International Handbook of Contemporary Research*. Berlin, de Gruyter. I, 506-569.
6. Culicover, Peter (1997). *Principles and Parameters. An Introduction to Syntactic Theory*. Oxford etc.: Oxford University Press.
7. Gelderen, Elly van (2003). The CP and split-CP cross-linguistically. Manuscript, Arizona State University.
8. Ghomeshi, Jila (1997). Topics in Persian VPs. *Lingua* 102, 133-167.
9. Grewendorf, Günther (2002). *Minimalistische Syntax*. Tübingen: UTB/Francke.
10. Hopper, Paul J. & Elizabeth Closs Traugott (1993, ²2003). *Grammaticalization*. Cambridge: Cambridge University Press.
11. Kaiser, Stefan & al. (2001). *Japanese: A Comprehensive Grammar*. New York: Routledge.
12. Karimi, Simin (2003). On Object Positions, Specifity and Scrambling in Persian. In: Simin Karimi (ed.), *Word Order and Scrambling*. Amsterdam: Benjamins. 91-125.
13. Kayne, Richard S. (1984). Connectedness. *Linguistic Inquiry* 14. 223-49.
14. Klaiman, Miriam (1976). *A Functional View on Some Syntactic Movement Typologies*. MA thesis, Univ. of Chicago.
15. Lazard, Gilbert (1992). *A Grammar of Contemporary Persian*. Costa Mesa etc.
16. Lightfoot, David (1979). *Principles of Diachronic Syntax*. Cambridge: CUP.
17. Lotfi, Ahmad R. (1998). Word Order and Accusative Case Assignment in Modern Persian. Paper presented at the 4th Linguistics Conference, Allameh Tabataba'i University, Tehran.
18. Meinunger, André (2004). Verb position, verbal mood, and the anchoring (potential) of sentences. In: Lohnstein, Horst & Susanne Trissler (eds.) . *The Syntax and Semantics of the Left Periphery*. Berlin, New York: Mouton de Gruyter. 313-42.
19. Müller, Gereon (1993). *On Deriving Movement Type Asymmetries*. Dissertation, Univ. Tübingen. (SfS-Report-05-93).

20. Müller, Gereon & Wolfgang Sternefeld (1993). Improper Movement and Unambiguous Binding. *Linguistic Inquiry* 24/3, 461-507.
21. Öhl, Peter (2003). *Economical Computation of Structural Descriptions in Natural Language*. PhD. Dissertation, University of Stuttgart.
22. Öhl, Peter & Agnes Korn (2006). Performative und parametrische Wandel in der linken Satzperipherie des Persischen. Der Subordinationsmarkierer *ke* und die Interrogativpartikel *āyā*. *Die Sprache* 46, 137-202.
23. Pesetsky, David (1982). *Paths and categories*. Cambridge: MIT Working Papers in Linguistics.
24. Pesetsky, David and Esther Torrego (2007). The syntax of valuation and the interpretability of features. In *Simin Karimi & al. (eds.), Phrasal and Clausal Architecture Syntactic derivation and interpretation in honor of Joseph E.monds*. Amsterdam: Benjamins. 262–294.
25. Poletto, Cecilia (2000). *The higher functional field: Evidence from Northern Italian dialects*. New York [u.a.]: Oxford Univ. Press.
26. Rizzi, Luigi (1997). The Fine Structure of the Left Periphery. In: L. Haegeman (Hg.). *Elements of Grammar: Handbook in Generative Syntax*. Dordrecht: Kluwer, 281-337.
27. Rizzi, Luigi (2001). On the Position 'Int(errogative)' in the Left Periphery of the Clause. In: Cinque, G. & C. Salvi (2001): *Current studies in Italian syntax: essays offered to Lorenzo Renzi*. Amsterdam: North Holland. 287-96.
28. Roussou, Anna (2000). On the left Periphery. Modal Particles and Complementisers. *Journal of Greek Linguistics* 1, 65-94.
29. Stowell, Tim (1981). *Origins of Phrase Structure*. Ph.D. Dissertation, MIT.
30. Vikner, Sten (1995). *Verb movement and expletive subjects in the Germanic languages*. New York, NY : Oxford.