

# Как да модернизираме исторически текст?

Стефан Герджиков, Стоян Михов

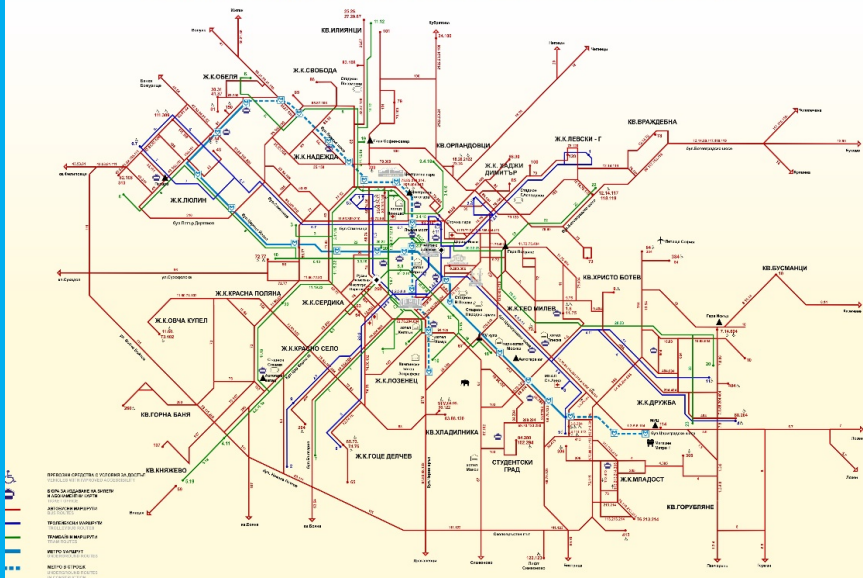
ФМИ Софийски университет, ИИКТ Българска академия на науките

2 Май 2012



# ЦЕНТЪР ЗА ГРАДСКА МОБИЛНОСТ SOFIA URBAN MOBILITY CENTRE

## СХЕМА НА МАСОВ ГРАДСКИ ТРАНСПОРТ В СОФИЯ SCHEME OF THE PUBLIC TRANSPORT IN SOFIA



мѣтнѡ, и искатъ отъ васъ основа що имъ сте длѣжни, и като не стесе още уготовили за плащанѣ. А ако. отъ друга страна, смислитесе за ваша должностъ, основа время. кое сначала се виждаше дълго, ще ви се вижда многѡ късо, колкѣто приближава; — ще да ви се почини, че прилѣпи и има крыла не самѡ на рамена си, но и на пѣты си, та за това толкѣва кѣрже пристига. Малки са велики пости законша кои иматъ да плащатъ должностъ на Воскресеніе, казѣва сиромахъ интѡнній.

Въ настоящее (сега), може, да слѣдѣватъ и да идатъ ваши работы добре, а слѣдственнѡ да имате на рѣка ги и довольны деньги (пары), и за това ви се чини че мало нещѡ по разслабѣванѣ во иждивеніа (харчовѣ) не може да въ повреди; но не е ли поразмно да чѣвате за старостъ и за зло время сичко нещѡ що имате повече? Утрешно слѣнце не дѣржи и не е секога сичкій день свѣтелъ. Придобыванѣ е небѣрно и привременно, но иждивеніа са нѣжни и истинни. По легко е, споредъ сиромаха интѡннія, да

и повече: Вой не брои за нищо заповѣди и совѣтѣванѣта, скорѡ се подлага на наказаніа и ударанѣта, отъ право слово и права рѣчь.

Така изговори старецъ Авраамъ. Сички слѣдѣваха неговъ разѣмны разговорки, и похвалиха ги; но веднага направиха все напротивъ, каквото става обыкновеннѡ и въ соборни побченіа, кои со съ вниманіе се слѣшатъ, а рѣдкѡ се праватъ; и секій отъ тѣхъ слѣшатели кѣпѣваше повече и преносходнѡ отъ своа си сила (кѣдрета). Колкѣто споредъ мене, азъ се увѣрихъ че той докѣ старецъ беше прочелъ внимателнѡ мои побчителны повседневици, и че кешесе занимавалъ и упражнявалъ прилѣжнѡ во основа, колкѣто кѣхъ говорилъ за таквѣя неща въ разстояніе двадесать и пять годинъ. Тѣа разѣмни говоренѣта собиралъ самъ отъ многѡ народы и отъ многѡ времена така, каквото онѣа, кои послѣ отъ жетварное сокиранѣ сокиратъ остатки. Какво и да е обаче, азъ като повторѣвахъ честѡ тѣа умства пользѣвахся довольнѡ, и рѣшихся, сир. совзехъ намѣреніе, да се използвамъ и да използвамъ на пользу



- FP 7, IP проект за трансформиране на печатното наследство на Европа в електронно достъпни ресурси.
- Целта е пълно интегриране на интелектуалното и културното наследство в съвременните информационни и комуникационни технологии.
- Бюджет: 16,5 милиона евро (12,1 милиона евро субсидия от ЕК)
- Участници: 26 национални и регионални библиотеки, научни звена и комерсиални участници, сред които IBM, Israel и Abbyy.



[www.cultura-strep.eu](http://www.cultura-strep.eu)

- FP 7, STREP проект за облекчаване на разбирането и изучаването на дигитализирани документи с историческа стойност.
- Целта е създаването на адаптивна изследователска среда, съобразена с потребителя, научната общност и характеристиките на изучаваните артефакти.
- Бюджет: 3,912 милиона евро (2,869 милиона евро субсидия от ЕК)
- Участници: Trinity College, Dublin; IBM, Israel; Софийски университет "Свети Климент Охридски"; Commetric Eood; University of Padua; Pintail Ltd; Technische Universität Graz.

fol. 290r

727

9o Aprilis 1642

The examinacion of Nicholas Woulfe of Clonefadue in the Quines Countie Joynener being duely sworne sworne & examined before vs afirmeth and sayth that on & since the Last of November 1641 there was taken away from him by the reblls of goods and Chattles moveable and vnmoveable to the vallue as alsoe due debts due vnto this deponent to the vallue of Two hundred pounds ster, the which goods were taken away from the Land aforesaid by Donnogh McDermott Grome of Cuddowe in the Countie aforesaid ag <A> together with Callagh Roe Walter Walsh of Ball Killbrittagh with divers others as this deponent wase informed by his servants haueing had one of his servants tyed by the rebells to a tree and threathned threatening that they the said rebbles would haue killed him this deponents servant Likewise this deponent deposeth that fflorence ffitz Patricke together with his consents threathned this deponent that vnlesse hee departed from his habitacion they would murther him And saith that the cause of their soe doeing was in regard this deponent went to the protestant Church and was dayly zealously bent against their religion and further sayth not that diuers rebells the said parties as abovesaid did have related that they are the Quines souldiers onely Nicholas Woolfe

Jurat die supradict

John Sterne

Randall Adams

вѣрѣванѣ, вѣрѣване → вярване

цѣль → цял, цел

совѣтъ → съвет

reblls, reblls, rebbles → rebels

ware → war, were

examinacion → examination

вѣрѣванѣ, вѣрѣване → вярване

цѣль → цял, цел

совѣтъ → съвет

reblls, reblls, rebbles → rebels

ware → war, were

examinacion → examination

Дадено:

$\mathcal{D}$  съвременен речник

$$\mathcal{I} = \{\langle h_i \rightarrow m_i \rangle\}_{i=1}^I$$

$$\mathcal{C} = \{\langle S_c \rightarrow T_c \rangle\}_{c=1}^C$$



вѣрѣванѣ, вѣрѣване → вярване

цѣль → цял, цел

совѣтъ → съвет

reblls, reblls, rebbles → rebels

ware → war, were

examinacion → examination

Дадено:

$\mathcal{D}$  съвременен речник

$$\mathcal{I} = \{\langle h_i \rightarrow m_i \rangle\}_{i=1}^I$$

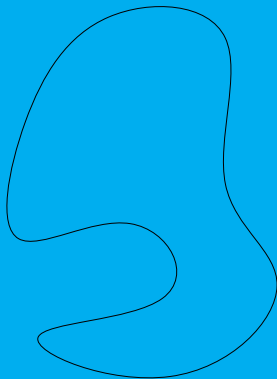
$$\mathcal{C} = \{\langle S_c \rightarrow T_c \rangle\}_{c=1}^C$$

Търси се: Модерен вариант на исторически текст  $S$ ?

Информация за	Модернизация	Превод
правописа на базовия език	Липса на норма	Фиксиран правопис
връзката между базовия и целевия език	частична (само примери $\mathcal{I}$ )	пълна
Словоред	сходен	различен

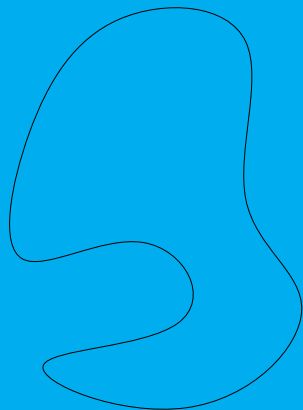
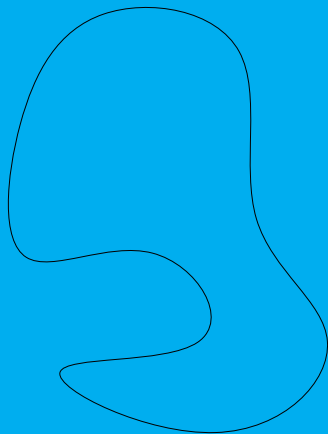
# Как да определим историческите вариации?

$$\mathcal{H} = \{h_i\}_{i=1}^I$$

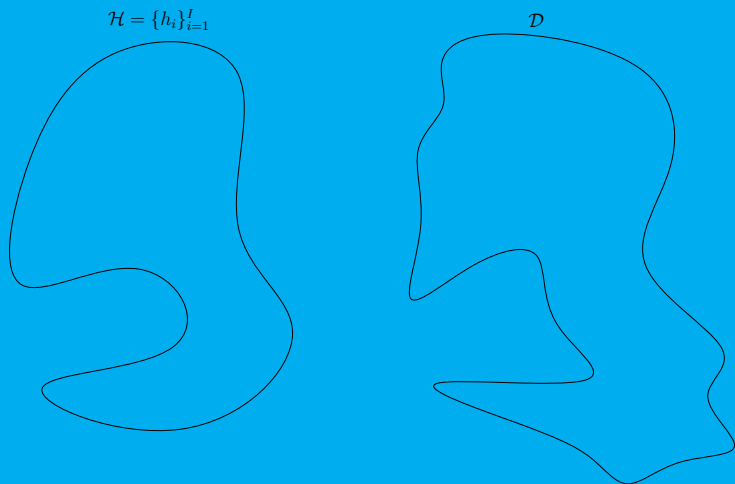


# Как да определим историческите вариации?

$$\mathcal{H} = \{h_i\}_{i=1}^I$$

 $\mathcal{D}$ 

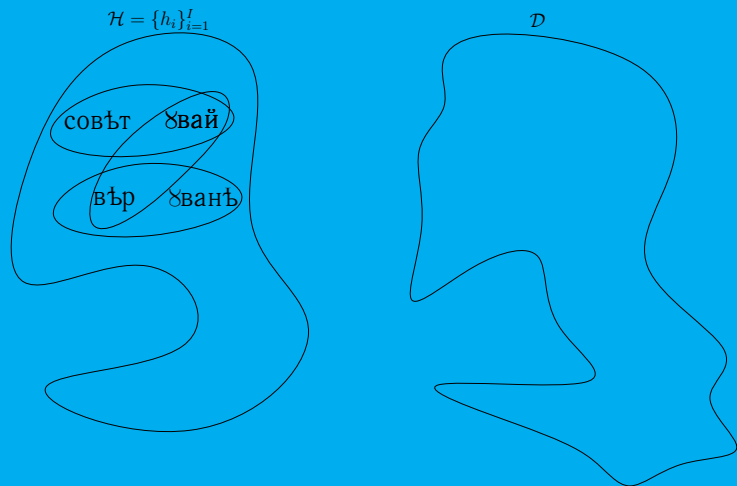
# Как да определим историческите вариации?



Blumer & Blumer et al, 1984, 1987:

$$\mathcal{S}_{\mathcal{H}} = \text{Pref}(\mathcal{H}) \cup \{\chi \mid \exists a \neq b \in \Sigma (a\chi, b\chi \in \text{Inf}(\mathcal{H}))\}$$

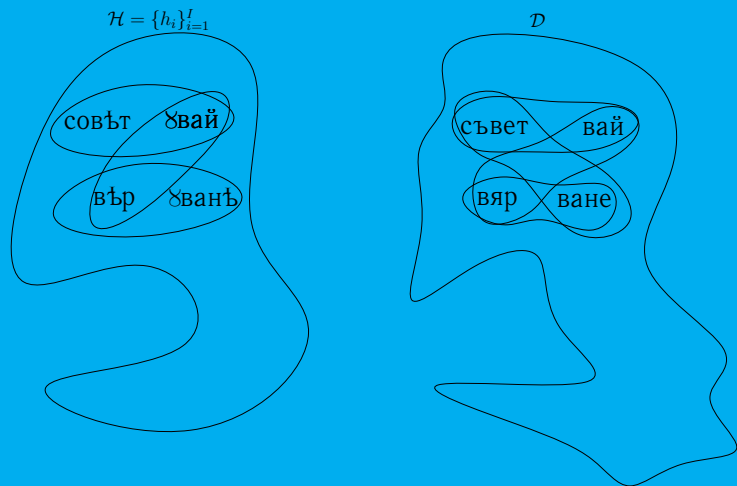
# Как да определим историческите вариации?



Blumer & Blumer et al, 1984, 1987:

$$\mathcal{S}_{\mathcal{H}} = \text{Pref}(\mathcal{H}) \cup \{\chi \mid \exists a \neq b \in \Sigma(a\chi, b\chi \in \text{Inf}(\mathcal{H}))\}$$

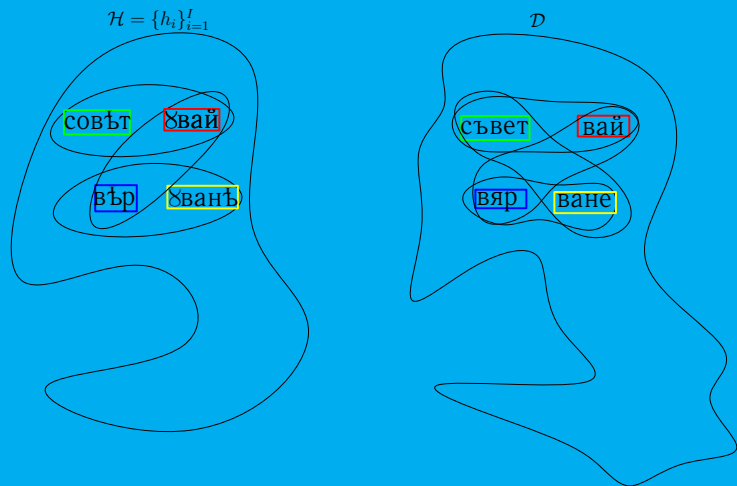
# Как да определим историческите вариации?



Blumer & Blumer et al, 1984, 1987:

$$\mathcal{S}_{\mathcal{H}} = \text{Pref}(\mathcal{H}) \cup \{\chi \mid \exists a \neq b \in \Sigma(a\chi, b\chi \in \text{Inf}(\mathcal{H}))\}$$

# Как да определим историческите вариации?

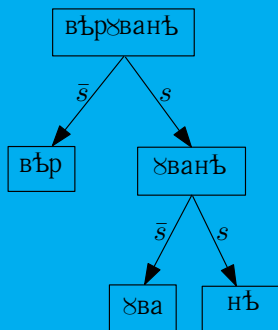


Blumer & Blumer et al, 1984, 1987:

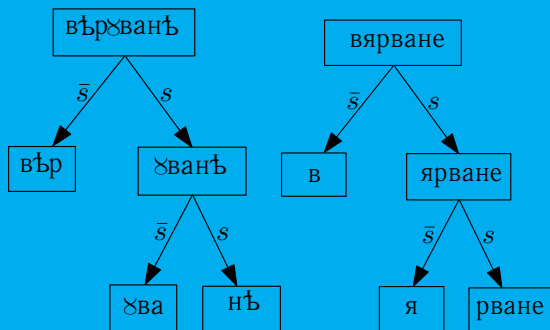
$$\mathcal{S}_{\mathcal{H}} = \text{Pref}(\mathcal{H}) \cup \{\chi \mid \exists a \neq b \in \Sigma(a\chi, b\chi \in \text{Inf}(\mathcal{H}))\}$$



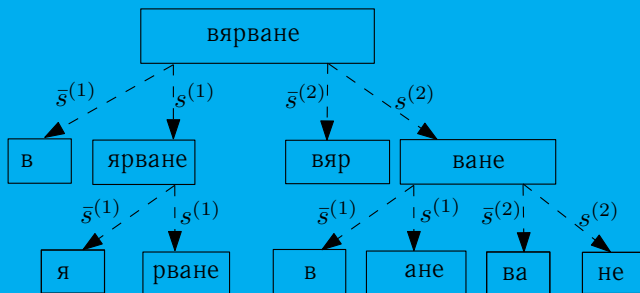
# Определяне на историческите вариации



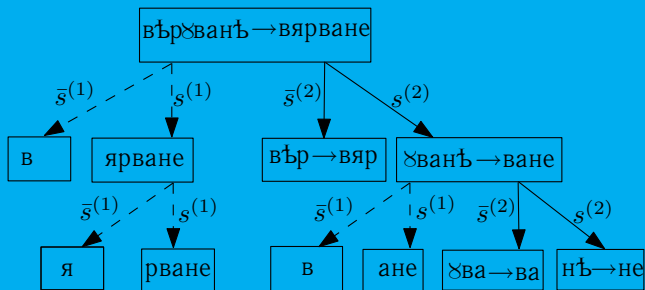
# Определяне на историческите вариации



# Определяне на историческите вариации



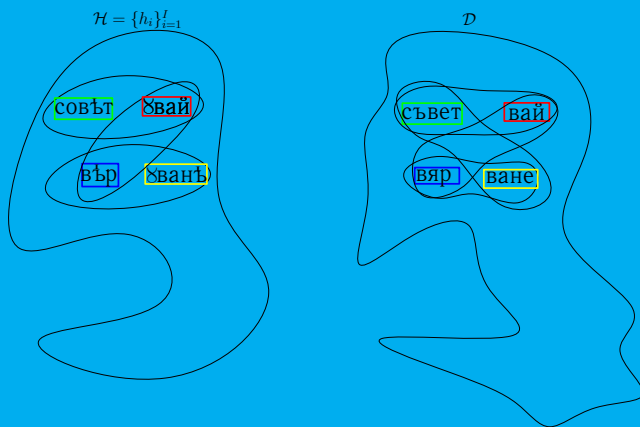
# Определяне на историческите вариации



$$suf(\chi, \mu) = \begin{cases} \max(|\chi|, |\mu|) - 1, & \chi^{rev} \wedge \mu^{rev} \neq \varepsilon \\ \max(|\chi|, |\mu|), & \chi^{rev} \wedge \mu^{rev} = \varepsilon \end{cases}$$

$$d_L(\chi, \mu) = \begin{cases} \max(|\chi|, |\mu|), & \chi = \varepsilon \in \{\chi, \mu\} \\ \min\{suf(\chi, \mu), \min_k d_L(\bar{s}(\chi), \bar{s}^{(k)}(\mu)) + d_L(s(\chi), s^{(k)}(\mu))\} \end{cases}$$

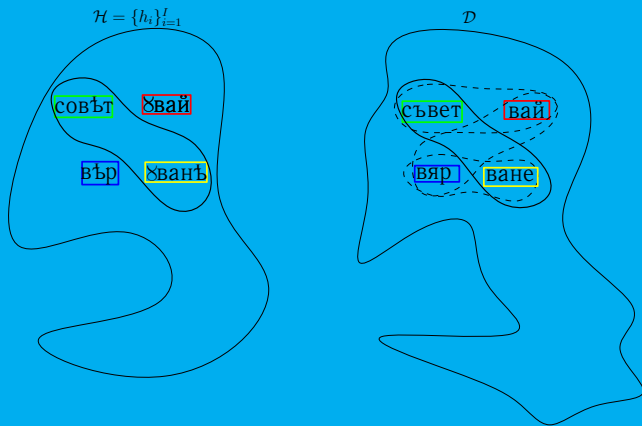
# СОВѢТЪВАНЪ



$$\mathcal{V} = \{ \langle \chi, \mu \rangle \mid \#(\chi, \mu) > 0 \}$$

$$p(\chi \rightarrow \mu) = \frac{\#(\chi, \mu)}{\sum_{\mu'} \#(\chi, \mu')}$$

# СОВѢТЪВАНЪ



$$\tilde{p}(h \rightarrow m) = \max \prod_{j=1}^n p(\chi_j \rightarrow \mu)$$

$$\chi_1 \circ \chi_2 \dots \chi_n = h \quad \chi_j \in \mathcal{T}_{\mathcal{H} \cup \{h\}}(h)$$

Och and Ney

The Alignment Template Approach to Statistical Machine Translation

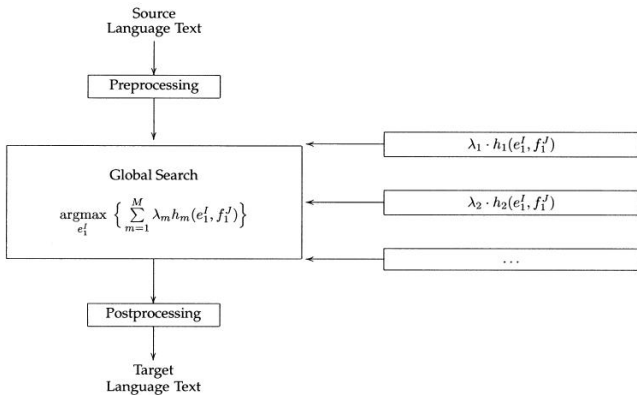
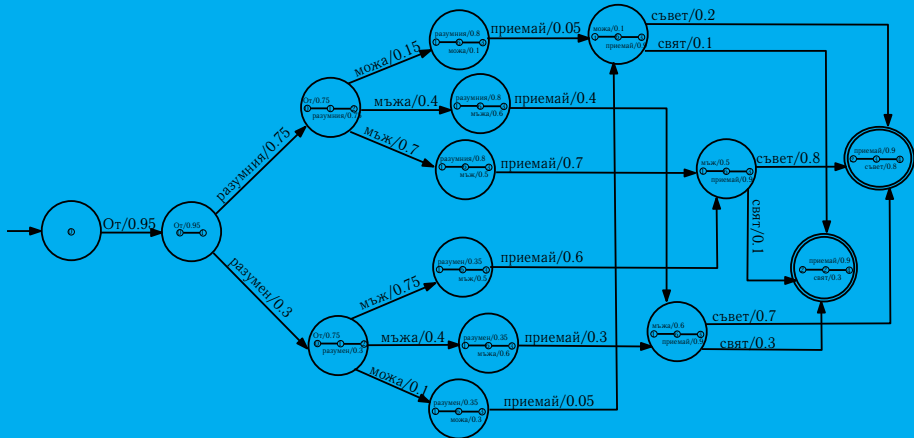


Figure 1

Architecture of the translation approach based on a log-linear modeling approach.

# Модернизация на исторически текст

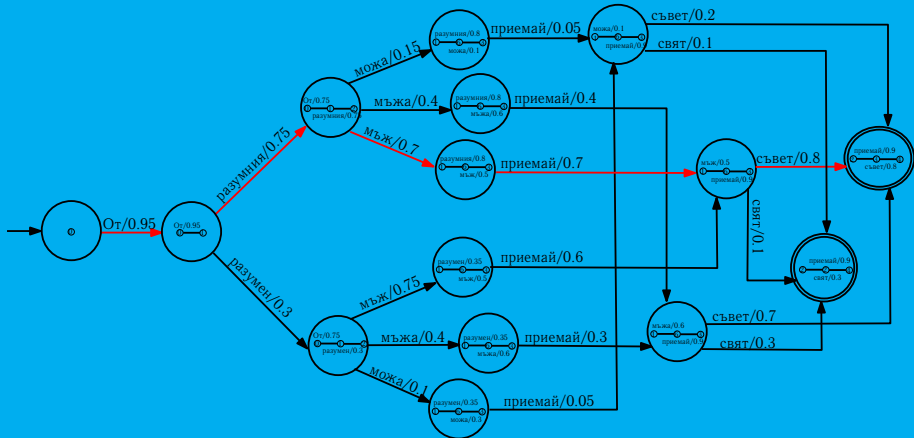
Отъ раздмнагѡ мъжа **приймай** совѣтъ.





# Модернизация на исторически текст

Отъ разѹмнаго мъжа **приймай** совѣтъ.





# ЦЕНТЪР ЗА ГРАДСКА МОБИЛНОСТ SOFIA URBAN MOBILITY CENTRE

## СХЕМА НА МАСОВ ГРАДСКИ ТРАНСПОРТ В СОФИЯ SCHEME OF THE PUBLIC TRANSPORT IN SOFIA

